

# Stochastic differential equation mixed effects models for tumor growth and response to treatment

Umberto Picchini<sup>a,\*</sup>, Julie Lyng Forman<sup>b,\*</sup>

<sup>a</sup>Centre for Mathematical Sciences, Lund University

Sölvegatan 18, SE-22100 Lund, Sweden

Email: [umberto@maths.lth.se](mailto:umberto@maths.lth.se)

<sup>b</sup>Section of Biostatistics, Department of Public Health, University of Copenhagen

Øster Farimagsgade 5, DK-1014 Copenhagen K, Denmark

Email: [jufo@biostat.ku.dk](mailto:jufo@biostat.ku.dk)

## Abstract

We model the growth dynamics for repeated measurements of tumor volumes in mice. We consider a two compartments representation corresponding to the fractions of tumor cells killed by and survived to a treatment, respectively. Dynamics are modelled with stochastic differential equations, resulting in a new state-space stochastic differential equation mixed effects model (SDEMEM) for response to treatment and regrowth. Inference for SDEMEMs is challenging due to the intractable likelihood function. We were able to estimate the model parameters, using both exact Bayesian methodology and an approximate method using the synthetic likelihoods approach. As a case study we consider data from two treatment groups and one control from a tumor xenography study, each consisting of 7-8 mice. Results from the case study and from a further simulation study shows that our models are able to reproduce the observed patterns and that Bayesian synthetic likelihoods is a reliable inference tool for SDEMEMs.

**Keywords:** Bayesian inference; repeated measurements; state-space model; particle MCMC; synthetic likelihood; tumor xenography.

## 1 Introduction

Preclinical cancer trials aim at understanding the dynamics of tumor growth and to evaluate the effect of treatments such as radio- and chemotherapies in delaying this. A typical trial involves repeated measurements of the volume of solid tumors grown in mice. Tumors are grown until a critical size is reached, in case of which the mouse must be sacrificed for ethical reasons, or until a planned end of study. Data from these trials pose a statistical challenge due to the missing data caused by the sacrifice and due to the substantial variation in growth patterns between subjects. Even within the same treatment group, it occurs that some tumors are eliminated following treatment, other continue to grow unaffected, and yet others display a decrease in volume followed by regrowth, [Laajala et al. \[2012\]](#).

Heitjan et al. [1993] review traditional approaches to analysing tumor xenography experiments, including ANOVA, MANOVA, and linear mixed models for tumor volumes. An overall drawback of the linear models is that inference target the mean log-tumor volume. This is problematic since tumor volumes are very often censored due to ethical guidelines that prohibit large size tumor volumes from ever being observed. An alternative comparison of two treatments can be obtained from a log-rank test on the time to sacrifice, tumor doubling times, or similar survival outcomes. Other approaches to analyzing tumor growth which rely on methods from survival analysis include models of tumor delay and log cell kill (inferred from tumor quadrupling times), see Stuschke et al. [1990], Wu and Houghton [2009] and Wu [2011]. However, these approaches have limited efficiency as they do not make use of the full information in the data. Moreover, delays and doubling times may in practice be hard to measure accurately due to day-to-day perturbations in growth, measurement error, and discrete time follow-up.

Demidenko [2013] reviews non-linear mixed models for tumor growth including the exponential, double exponential and delayed double exponential model for re-growth following treatment. We will focus on the double exponential model which identifies two latent compartments corresponding to the fraction of the tumor which is killed by the treatment and the one that survives. This model offers a phenomenological explanation for the variation in individual tumor growth patterns, by recognizing a) the proportion of the tumor killed by treatment, b) the rate of elimination of the dead tumor cells, and c) the growth rate of the surviving part of the tumor. Clinically relevant quantities such as tumor doubling times, tumor growth delay and surviving fraction of tumor cells can be deduced from the double exponential model, as shown in Demidenko [2006] and Demidenko [2010]. More recent non-linear mixed model approaches specify the individual growth curves semi-parametrically [Xia et al., 2013], or as splines (Kong and Yan, 2011, Zhao et al., 2011). These models allow for much flexibility in individual growth curves but do not share the biological interpretation of the double exponential or delayed double exponential models. A drawback of the double exponential, and other classical non-linear mixed models, is that the only source of intra-subject variation is given by independent identically distributed measurement errors. This is not realistic as growth rates are subject to day-to-day variation, due to biological processes not easily accounted for.

In recent years, a number of works have promoted the use of stochastic differential equation mixed effects models (SDEMEmS) as a more realistic alternative to the classical nonlinear mixed models. For instance, Donnet et al. [2010] find that a stochastic differential equation version of the Gompertz growth model is superior to its nonlinear deterministic mixed model counterpart for prediction of the body weight of growing chicken. Donnet and Samson [2013] report similar findings from pharmacokinetic experiments. Recent contribution to SDEMEmS inference are Delattre and Lavielle [2013] and Whitaker et al. [2016], see in particular the latter for an account on recent contributions. In general, even when not considering experiments involving repeated experiments, inference for stochastic differential equations (SDEs) is challenging [Fuchs, 2013], because nonlinear SDE models have unknown transition densities, hence intractable likelihood functions. An additional difficulty is that we do not assume availability of measurements from a Markov process (solution to an SDE), instead we consider measurements to be affected by measurement noise. Hence the class of models we formulate in this paper is of state-space type (hidden Markov models, Cappé et al., 2006). It is possible to treat state-space models, including those having dynamics given by an SDE, by employing

state-of-art likelihood-based methods using approximations given by sequential Monte Carlo (SMC) filters. Indeed, in one of our attempts at estimating model parameters for our SDEMEMs, we use a particle marginal method [Andrieu and Roberts, 2009] which returns exact Bayesian inference, despite employing SMC approximations. However we also consider a methodology which is able to target more general models, beyond the state space class: we use a Bayesian version of the synthetic likelihoods (SL) approach due to Price et al. [2016]. SL was initially proposed in Wood [2010] and does not impose any assumption on the complexity of the model, the only requirement being the ability to simulate artificial datasets from the model and construct summary statistics for the generated datasets. By producing a comparison with the exact Bayesian parameter inference resulting from the particle marginal method, we show that reliable inference using Bayesian SL (BSL) is possible for SDEMEMs, both with real data and with artificial data produced in a simulation study. Interestingly, for the simulation study we show that point estimates (posterior means) from BSL have a smaller bias than those from exact Bayesian inference. This might point to a finding in Wood [2010], where in the context of nearly chaotic ecological models he found that replacing the raw data with corresponding summary statistics can make parameter inference more robust to sources of randomness in stochastic modelling. Indeed SDEMEMs are complex models with three layers of randomness, as described in next sections. Therefore SL can be considered as an additional tool for inference in this class of models. Most importantly, in future studies we could redefine the model to be of non-state-space type, and in such scenario exact parameter inference via sequential Monte Carlo (particle) methods might not be an option, while SL can still be applied. We also construct a stochastic EM algorithm for maximum likelihood estimation, with scarce success.

The structure of the paper is as follows: Section 2 contrasts the classical nonlinear mixed models with the stochastic differential equation mixed effects models (SDEMEMs). Section 3 considers a particle marginal method for exact Bayesian inference in SDEMEMs. Section 4.2 considers the BLS approach using synthetic likelihoods. To the best of our knowledge this is the first application of synthetic likelihoods to SDEMEMs. We show that the simple approach offered by the synthetic likelihoods methodology is effective for the investigated SDEMEMs. In section 5 we illustrate the construction of a maximum likelihood strategy using a stochastic version of the EM algorithm: the analytic construction of the procedure is quite lengthy and its practical implementation certainly not trivial. We were not able to obtain reasonable maximum likelihood inference, though a discussion on a possible solution is carried out in section 6.3. In section 6 we analyse data from a tumor xenography study including two treatment groups and a control each consisting of 7-8 mice. Results indicate that the suggested models are appropriate for describing tumor growth and response to treatment, but some model parameters are poorly identified due to the small sample size. Finally, in section 7 we run two simulation studies on artificial data. The first simulations shows that Bayesian synthetic likelihoods is a reliable inference approach for SDEMEMs, when compared to results from an exact Bayesian algorithm. The second simulations shows that with a larger number of subjects our model is able to distinguish the treatment efficacy between two different treatment groups.

## 2 Mixed effects models of tumor growth

### 2.1 Ordinary mixed effects models

Denote with  $M$  the number of subjects in a given treatment group. Assume that tumor volumes from subject  $i$  are measured at time points  $t_{i0} < \dots < t_{in_i}$ ,  $i = 1, \dots, M$ . In a planned experiment, such as the one considered in section 6, time points will usually be the same for all subjects, i.e.  $t_{ij} = t_j$ , but the number of observations  $n_i$  may differ between subjects. For instance, the mice in the above mentioned experiment were sacrificed when their tumor volume exceeded a critical size prescribed by the ethical guidelines. Denote with  $V_{ij} = V_i(t_j)$  the exact tumor volume of subject  $i$  at time  $t_j$ ,  $j = 1, \dots, n_i$ . We model the observations as

$$Y_{ij} = \log(V_{ij}) + \varepsilon_{ij}, \quad i = 1, \dots, M; j = 1, \dots, n_i \quad (1)$$

where the  $\varepsilon_{ij}$ 's are i.i.d. normally distributed measurements errors with  $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ . This means that we assume tumor volumes to be measured with multiplicative log-normal measurement errors.

In regulated experiments the mice are sacrificed long before the tumor volumes reach steady state. Hence, unperturbed growth in the control group is adequately described by a simple exponential growth model. Let  $\beta_1, \dots, \beta_M$  denote the subject specific growth rates, then the growth curves for the control group are given by

$$\frac{dV_i(t)}{dt} = \beta_i V_i(t), \quad V_i(0) = v_{i,0}, \quad i = 1, \dots, M. \quad (2)$$

Of course, this is solved explicitly by  $V_i(t) = v_{i,0}e^{\beta_i t}$ . Note that with the further assumption that growth rates are normally distributed and initial tumor volumes log-normally distributed across the population, the observation model (1) is merely a standard linear mixed model with a random intercept and a random slope.

If tumor volumes are observed post treatment, then the double exponential model in [Demidenko \[2013\]](#) describes the total volume in terms of surviving tumor cells and cells killed by the treatment as

$$\begin{aligned} V_i(t) &= V_i^{\text{surv}}(t) + V_i^{\text{kill}}(t), \\ \frac{dV_i^{\text{surv}}(t)}{dt} &= \beta_i V_i^{\text{surv}}(t), \quad V_i^{\text{surv}}(0) = (1 - \alpha_i)v_{i,0}, \quad i = 1, \dots, M \\ \frac{dV_i^{\text{kill}}(t)}{dt} &= -\delta_i V_i^{\text{kill}}(t), \quad V_i^{\text{kill}}(0) = \alpha_i v_{i,0} \end{aligned} \quad (3)$$

where  $\alpha_i \in [0, 1]$  denotes the proportion of the tumor that has been killed by the treatment and  $\delta_i$  denotes the elimination rate for the dead tumor cells. Equation (3) has the explicit solution  $V_i(t) = (1 - \alpha_i)v_{i,0}e^{\beta_i t} + \alpha_i v_{i,0}e^{-\delta_i t}$ . Note that if tumors were allowed to grow beyond the limit of the ethical guidelines, then the volumes of long-term surviving mice would eventually reach a steady state. In this case data would more adequately be described by growth curves such as the Gompertz, Richards, Weibull or Logistic, see e.g. [Heitjan \[1991\]](#).

### 2.2 Stochastic differential equation mixed effects model

The assumption of time constant growth and elimination rates in the ordinary mixed models is usually not realistic, since growth is affected by various biological processes that

are not easily accounted for. [Donnet et al. \[2010\]](#) used a stochastic differential equation (SDE) version of the Gompertz curve to model the body weights of chicken over time. They found that the SDE model provided much more accurate dynamical predictions of individual weights, compared to ordinary differential equation models.

We therefore suggest to replace the ordinary differential equation model specified by (2) with a SDE model such as the geometric Brownian motion,

$$dV_i(t) = (\beta_i + \gamma^2/2)V_i(t)dt + \gamma V_i(t)dB_i(t), \quad V_{i,0} = v_{i,0}. \quad (4)$$

Here the  $\{B_{i,t}\}_{t \geq 0}$ 's are independent standard Brownian motions and  $\gamma^2$  denotes the intra-subject growth rate variance. This means that the instantaneous growth rate is not exactly  $\beta_i$  but deviates from this by a random normal perturbation (white noise). The motivation for including the term  $\gamma^2/2$  in the drift of the SDE is that the individual growth process is then given by  $V_i(t) = v_{i,0}e^{\beta_i t + \gamma B_i(t)}$  which is a log-normally distributed stochastic process with median (and geometric mean)  $v_{i,0}e^{\beta_i t}$ , which coincides with the ordinary exponential growth model (2). With the further assumption that growth rates are distributed as  $\beta_i \sim \mathcal{N}(\bar{\beta}, \sigma_\beta^2)$  and initial tumor volumes as  $\log(v_{i,0}) \sim \mathcal{N}(\bar{v}_0, \sigma_0^2)$  across the population, then it follows that volumes at time  $t_{ij}$  follow a log-normal distribution with median (and geometric mean)  $\bar{v}_0 e^{\bar{\beta} t_{ij}}$  which is the same as in the ordinary log-linear mixed model.

Similarly, the ordinary double exponential model (3) can be replaced by a stochastic differential equation mixed effects model with the following specification

$$\begin{aligned} Y_{ij} &= \log(V_{ij}) + \varepsilon_{ij}, & i = 1, \dots, M; \quad j = 1, \dots, n_i \\ V_i(t) &= V_i^{\text{surv}}(t) + V_i^{\text{kill}}(t), \\ dV_i^{\text{surv}}(t) &= (\beta_i + \gamma^2/2)V_i^{\text{surv}}(t)dt + \gamma V_i^{\text{surv}}(t)dB_i(t), & V_i^{\text{surv}}(0) = (1 - \alpha_i)v_{i,0} \\ dV_i^{\text{kill}}(t) &= (-\delta_i + \tau^2/2)V_i^{\text{kill}}(t)dt + \tau V_i^{\text{kill}}(t)dW_i(t), & V_i^{\text{kill}}(0) = \alpha_i v_{i,0}. \end{aligned} \quad (5)$$

The  $\{W_i(t)\}_{t \geq 0}$ 's are additional standard Brownian motions assumed mutually independent and independent of the  $\{B_i(t)\}_{t \geq 0}$ 's, of the  $\varepsilon_{ij}$  and the system initial conditions. Here  $\tau^2$  denotes the intra-subject elimination rate variance.

Finally, we point out that models (4) and (5) by no means are the only possible ways of specifying SDEMEmS for tumor growth. Diffusion terms of the type  $\gamma V_i(t)dB_i(t)$  could be replaced by a more flexible form such as  $\gamma V_i(t)^\rho dB_i(t)$  (for some exponent  $\rho$ ), or by the diffusion term resulting from assuming particular dynamics for the time-varying growth rate, e.g. the Ornstein-Uhlenbeck process  $d\beta_i(t) = -\rho(\beta_i(t) - \bar{\beta})dt + \gamma dB_i(t)$ . More general models such as these do not admit simple closed form solutions like (4) and (5) do, but can still be analyzed using the Bayesian methods described in sections 3 and 4.2 below.

### 3 Likelihood-based inference for SDEMEmS

In this section we discuss likelihood inference for SDEMEmS such as model (4) and (5) and generalizations thereof. Notice that all models discussed in section 2.2 can be viewed as instances of the general state-space SDEMEm

$$\begin{cases} Y_{ij} &= g(X_i(t_{ij}), \varepsilon_{ij}), & \varepsilon_{ij} \sim_{i.i.d.} N(0, \sigma_\varepsilon^2) \\ dX_i(t) &= \mu(X_{it}, t, \phi_i)dt + \sigma(X_{it}, t, \phi_i, \kappa)dB_i(t), & X_i(t_0) \sim \pi_0(x_i(t_0)|\phi_i) \\ \phi_i &\sim_{i.i.d.} p(\phi_i|\eta). \end{cases} \quad (6)$$

Here each process  $\{X_i(t)\} \in \mathcal{X}$  has dimension  $d_x$  and is defined on the space  $\mathcal{X}$ . Each  $Y_{ij} \in \mathcal{Y}$  has dimension  $d_y$  and is defined on the space  $\mathcal{Y}$ . Model (5) is a specific instance of (6) having  $d_x = d_y = 1$ . Model (6) has the following interpretation: for each subject  $i$ ,  $\{X_{it}\}_{t \geq 0}$  represents the state of the hidden (unobservable) biological process of interest at time  $t$ . Direct observation of  $\{X_i(t)\}$  is assumed impossible: instead we denote with  $Y_{ij}$  a generic observation collected at discrete time  $t_{ij}$ , this being subject to measurement error modeled via the known function  $g(\cdot)$ , with additive measurement error  $g(x, \varepsilon) = x + \varepsilon$  and multiplicative measurement error  $g(x, \varepsilon) = e^\varepsilon \cdot x$  being the most common specifications. The dynamics of the latent process are modeled via the drift and diffusion functions  $\mu(\cdot)$  and  $\sigma(\cdot)$  which are assumed known, save from model parameters including the subject specific random effect  $\phi_i$ . Each individual random effect  $\phi_i$  is assumed distributed with a density  $p(\cdot|\beta)$  depending on the “population parameter”  $\beta$ , the latter being common to all subjects. Conditionally on  $\phi_i$ , the solution process  $\{X_i(t)\}$  to (6) for each  $i$  is subject to the same regularity conditions for the existence and uniqueness of the solution of an SDE; such conditions can be found e.g. in [Fuchs \[2013\]](#). We assume that measurements  $Y_{ij}$  are conditionally independent given the latent states  $X_{ij} := X_i(t_j)$ , implying that model (6) is a state space model [[Cappé et al., 2006](#)]. However notice that while Markovianity of the latent states  $\{X_i(t)\}$  and conditional independence of measurements are essential when dealing with inference methods described in the present section, these are not required properties when employing the methodology in section 4.2. Goal of our study is to perform inference for the vector parameter  $\theta = (\eta, \kappa, \sigma_\varepsilon)$ . One of the interesting features of the SDEM (6) is its ability to discriminate between the intra-subject variability (the parameter  $\kappa$ ), the inter-subjects variability (the variance of the individual parameter  $\phi_i$ ) and the residual variability ( $\sigma_\varepsilon$ ). Therefore each measurement  $Y_{ij}$  depends on three sources of variability: the variability of the individual stochastic dynamics  $\{X_i(t)\}$ , the specific realization of  $\phi_i$  and the residual error  $\varepsilon_{ij}$ .

Denote by  $y_i = \{y_{ij}\}_{j=1, \dots, n_i}$  the observations for subject  $i$  and by  $X_i = \{X_{ij}\}_{j=1, \dots, n_i}$  the corresponding values of the latent process. Let  $y = (y_1, \dots, y_M)$  denote the full dataset containing measurements for all subjects in the considered group. Standard methods for frequentist as well as Bayesian estimation of the model parameters  $\theta = (\eta, \kappa, \sigma_\varepsilon)$  require the evaluation of the likelihood function  $p(y|\theta) = \prod_{i=1}^M p(y_i|\theta)$ . The hidden Markov structure implies the following derivation

$$\begin{aligned}
p(y_i|\theta) &= \int p(y_i|\phi_i; \theta) p(\phi_i|\theta) d\phi_i \\
&= \int \left( \int p(y_i|X_i; \theta) p(X_i|\phi_i; \theta) dX_i \right) p(\phi_i|\theta) d\phi_i \\
&= \int \left( \int \left\{ \prod_{j=1}^{n_i} p(y_{ij}|X_{ij}, \phi_i, \theta) p(X_{i,j}|X_{i,j-1}, \phi_i; \theta) \right\} p(X_{i0}|\phi_i, \theta) dX_i \right) p(\phi_i|\theta) d\phi_i.
\end{aligned} \tag{7}$$

Note that the term  $p(X_{i0}|\phi_i, \theta)$  vanish in either of the cases where  $X_{i0}$  is included among the random effects or is assumed to be a known constant,  $X_{i0} := x_{i0}$ .

In general the likelihood function (7) is not analytically tractable. Thus, inference for SDEMEMs rely on either more specific model assumptions such as the latent processes being Gaussian, or the use of computationally intensive Bayesian methods. [Delattre and Lavielle \[2013\]](#) show how to conduct likelihood inference in SDEMEMs by use of the stochastic approximate EM algorithm (SAEM) coupled with an extended Kalman filter. [Donnet and Samson \[2014\]](#) propose to use a particle MCMC algorithm to conduct



the S-step in SAEM. In either case the use of SAEM requires explicit specification of sufficient summary statistics for the augmented likelihood  $p(y_i, X_i, \phi_i | \theta)$ . While providing fast and accurate inference in models with a latent Gaussian structure, the derivation of the summary statistics is a tedious if not impossible task for more complex models of realistic interest, see an example in section 5.1. See also Picchini [2016] for a likelihood-free version of SAEM.

Bayesian inference targets the posterior distribution  $\pi(\theta|y) \propto p(y|\theta)\pi(\theta)$  where  $\pi(\theta)$  is the prior distribution. Bayesian methodology for SDEMEmS was first studied by Donnet et al. [2010] who implemented a Gibbs sampler that applies to the case where the SDE has an explicit solution, and which can be extended to the more general state space model by using an Euler-Maruyama discretization. A recent review of Bayesian inference methods for SDEMEmS can be found in Whitaker et al. [2016]. It is important to notice that MCMC algorithms can be constructed to sample from the exact posterior of model (6). This was first noted by Beaumont [2003] and later formalized by Andrieu and Roberts [2009] in the context of general state space models.

Here we exemplify the *pseudo-marginal* particle method [Andrieu and Roberts, 2009], which uses a sequential Monte Carlo (SMC) filter to approximate the intractable likelihood function (7). The key idea is to substitute the intractable likelihood  $p(y|\theta)$  by an unbiased non-negative estimate  $\hat{p}(y|\theta)$  in an otherwise standard Metropolis-Hastings algorithm (see algorithm 1 below).

---

**Algorithm 1** A pseudo-marginal MCMC algorithm

---

1. **Input:** a positive integer  $R$ . Fix a starting value  $\theta^*$  or generate it from its prior  $\pi(\theta)$  and set  $\theta_1 = \theta^*$ . Set a kernel  $q(\theta'|\theta)$ . Use algorithm 2 to obtain an unbiased estimate  $\hat{p}(y|\theta^*)$  of  $p(y|\theta^*)$ .

**Output:**  $R$  correlated draws from  $\pi(\theta|y)$  (possibly after a burnin).

2. Generate a  $\theta^\# \sim q(\theta^\#|\theta^*)$ . Use algorithm 2 to obtain an unbiased estimate  $\hat{p}(y|\theta^\#)$  of  $p(y|\theta^\#)$ .

3. Generate a uniform random draw  $u \sim U(0, 1)$ , and calculate the acceptance probability

$$\alpha = \min \left[ 1, \frac{\hat{p}(y|\theta^\#)}{\hat{p}(y|\theta^*)} \times \frac{q(\theta^*|\theta^\#)}{q(\theta^\#|\theta^*)} \times \frac{\pi(\theta^\#)}{\pi(\theta^*)} \right].$$

If  $u > \alpha$ , set  $\theta_{r+1} := \theta_r$  otherwise set  $\theta_{r+1} := \theta^\#$ ,  $\theta^* := \theta^\#$  and  $\hat{p}(y|\theta^*) := \hat{p}(y|\theta^\#)$ . Set  $r := r + 1$  and go to step 4.

4. Repeat steps 2–3 as long as  $r \leq R$ .

---

Note that in sections 6–7 we used a Gaussian kernel  $q(\cdot|\cdot)$  to propose parameters via the adaptive Gaussian random walk algorithm of Haario et al. [2001].

An unbiased SMC-based estimate of the likelihood function is given by

$$\hat{p}(y|\theta) = \prod_{i=1}^M \hat{p}(y_i|\theta), \quad (8)$$

where for each  $i = 1, \dots, M$

$$\hat{p}(y_i|\theta) = \hat{p}(y_{i1}|\theta) \prod_{j=2}^{n_i} \hat{p}(y_{ij}|y_{i,1:j-1}, \theta) = \prod_{j=1}^{n_i} \left( \frac{1}{L} \sum_{l=1}^L w_{ij}^l \right), \quad (9)$$

$L$  is the number of particles and the  $w_{ij}^l$ 's are the *importance weights* obtained from the particle filter described in algorithm 2 below.

Algorithm 2 is an instance of the bootstrap filter algorithm [Gordon et al., 1993]. Note that it is merely for ease of presentation that algorithm 2 devise resampling at each time point (i.e. for each  $j$ ). In the case study of section 6 we resample only at time points where the effective sample size (ESS) is smaller than  $L/3$ , where an estimate of ESS is given by  $(\sum_{l=1}^L (\tilde{w}_{ij}^l)^2)^{-1}$  [Liu, 1996]. Moreover we performed the resampling step using the stratified resampling method of Kitagawa [1996]. We have not considered strategies to tune the value of  $L$ , however the reader can refer to Doucet et al. [2015] and Sherlock et al. [2015].

---

**Algorithm 2** SMC for mixed-effects state-space models

---

**Input:** a positive integer  $L$ , a starting value for  $\theta$  and a starting value  $x_0$ . Set time  $t_0 = 0$  and corresponding starting states  $X_{i0} = x_{i0}$ . We use the convention that all steps involving the index  $l$  must be performed for all  $l \in \{1, \dots, L\}$ .

**Output:** all the  $\hat{p}(y_{ij}|y_{i,1:j-1})$ ,  $i = 1, \dots, M$ ;  $j = 1, \dots, n_i$ .

**for**  $i = 1, \dots, M$  **do**

draw  $\phi_i^l \sim p(\phi_i|\theta)$

**if**  $j = 1$  **then**

Sample  $x_{i1}^l \sim p(x_{i1}|x_{i0}, \phi_i^l; \theta)$ .

Compute  $w_{i1}^l = p(y_{i1}|x_{i1}^l)$  and  $\hat{p}(y_{i1}) = \sum_{l=1}^L w_{i1}^l / L$ .

Normalization:  $\tilde{w}_{i1}^l := w_{i1}^l / \sum_{l=1}^L w_{i1}^l$ . Interpret  $\tilde{w}_{i1}^l$  as a probability associated to  $x_{i1}^l$ .

Resampling: sample  $L$  times with replacement from the probability distribution  $\{x_{i1}^l, \tilde{w}_{i1}^l\}$ . Denote the sampled particles with  $\tilde{x}_{i1}^l$ .

**end if**

**for**  $j = 2, \dots, n_i$  **do**

Forward propagation: sample  $x_{ij}^l \sim p(x_{ij}|\tilde{x}_{i,j-1}^l, \phi_i^l; \theta)$ .

Compute  $w_{ij}^l = p(y_{ij}|x_{ij}^l)$  and normalise  $\tilde{w}_{ij}^l := w_{ij}^l / \sum_{l=1}^L w_{ij}^l$

Compute  $\hat{p}(y_{ij}|y_{i,1:j-1}) = \sum_{l=1}^L w_{ij}^l / L$

Resample  $L$  times with replacement from  $\{x_{ij}^l, \tilde{w}_{ij}^l\}$ . Sampled particles are  $\tilde{x}_{ij}^l$ .

**end for**

**end for**

---

The main distinction between model (6) and other state space models is that latent states are subject specific and can be further decomposed into a time-dependent component  $X_i$  and a time-independent component  $\phi_i$ . Therefore, when applying SMC we first draw  $\phi_i$  and then, conditionally on such draw, we propagate forward particles corresponding to the states  $X_i$ . As proven by Del Moral [2004] and Pitt et al. [2012] each individual estimate (9) produced by the bootstrap filter is unbiased (where the expectation is taken with respect to the distribution used to generate the SMC approximation). As subjects are independent, it follows that  $\hat{p}(y|\theta)$  is an unbiased estimator for  $p(y|\theta)$  regardless of the choice of  $L$ . As shown in Beaumont [2003] and Andrieu and Roberts [2009], using an unbiased estimator for  $p(y|\theta)$  is sufficient to ensure that draws from algorithm 1 have stationary distribution  $\pi(\theta|y)$ .

## 4 Approximate inference for SDEMEmS

In this section we discuss approximate inference for SDEMEmS. We briefly discuss approximate Bayesian computation (ABC) and synthetic likelihoods, these being the most



popular methods for making inference in models with an intractable likelihood function. Common to both methodologies is that they do not require any evaluation of the likelihood function but solely rely on simulations from the model in consideration. It is important to notice that both methods do not require the model to have a state-space representation such as (6). For instance, both ABC and synthetic likelihoods would apply to a model without assuming Markovian dynamics or conditionally independent measurements, as long as realizations from this model can be simulated. Compared to exact likelihood-based methods, the drawbacks of approximate inference is the loss of statistical and computational efficiency and a need to validate the performance of the estimators on a case to case basis.

## 4.1 Approximate Bayesian Computation

Over the past ten years approximate Bayesian computation (ABC) has become increasingly popular for conducting inference in complex models within diverse fields of application. Here we only outline the key ideas around ABC and refer to [Sisson and Fan \[2011\]](#) and [Marin et al. \[2012\]](#) for reviews. ABC bases inference on a set of carefully selected set of summary statistics which are used in computation as a surrogate for the entire data set. In what follows we denote the summary statistics by  $s := s(y)$  which is assumed to be a  $d$ -dimensional vector. Algorithm 3 illustrates ABC in its simplest form, where parameters  $\theta$  are sampled from an approximation to the posterior distribution  $\pi(\theta|y)$  using an accept/reject step.

---

### Algorithm 3 Approximate Bayesian Computation

---

**Input:** a positive integer  $R$  and the observed summary statistics  $s$ . Set a tolerance  $\epsilon > 0$  and a distance measure  $\|\cdot\|$ .

**Output:**  $R$  samples from the approximate posterior  $\pi_\epsilon(\theta|s) := \pi(\theta \mid \|s(y^\#) - s\| < \epsilon)$ .

1. Sample  $\theta^\# \sim \pi(\theta^\#)$ . Conditionally on  $\theta^\#$  generate a dataset  $y^\# \sim p(y|\theta^\#)$ .
  2. Compute the summary statistic  $s^\# = s(y^\#)$  and the distance  $\text{dist}^\# = \|s^\# - s\|$ . If  $\text{dist}^\# < \epsilon$ , set  $\theta_r := \theta^\#$  and  $r := r + 1$  and go to step 3, otherwise go to step 1.
  3. Repeat steps 1–2 as long as  $r \leq R$ .
- 

Note that algorithm 3 accepts only those values of  $\theta$  that generate summary statistics sufficiently close to the observed  $s(y)$  in terms of the distance measure  $\|\cdot\|$  and the tolerance  $\epsilon$ . In an ideal scenario, the summary statistic  $s$  is sufficient for  $\theta$ , in which case ABC approximates the posterior distribution,  $\pi(\theta|y) = \pi(\theta|s)$  as  $\epsilon \rightarrow 0$ . In non-trivial applications, no set of summary statistics can ever be expected to be sufficient and simulation studies are thus needed to validate the performance of the ABC estimates. Besides the simple algorithm 3, many variants of ABC exist including MCMC ([Marjoram et al. \[2003\]](#)) and SMC hybrids ([Toni et al. \[2009\]](#)). In general, ABC algorithms are difficult to tune and their performance is very sensitive to the choice of  $\epsilon$ . Moreover, it is important to weight the distance measure in proportion to the variability of the individual summary statistics, otherwise a small  $\epsilon$  will only produce accurate inference for the parameter associated with the most variable summary in  $s$ , see [Prangle \[2015\]](#).

## 4.2 Synthetic likelihoods

The synthetic likelihoods (SL) methodology was introduced in [Wood \[2010\]](#) as a mean to produce inference in models with an intractable likelihood function and/or near-chaotic

dynamics. Similarly to ABC, SL relies on a set of carefully selected summary statistics. However, while in ABC no assumption is set on the distribution of  $s$ , with SL summary statistics are assumed to have a joint normal distribution,  $s \sim \mathcal{N}(\mu(\theta), \Sigma(\theta))$ . If this holds true, and if parameters in  $\theta$  can be identified from  $\mu(\theta)$  and  $\Sigma(\theta)$ , then inference for  $\theta$  can be based on the normal likelihood of  $s$  instead of the intractable likelihood of  $y$ . In most applications of SL the functions  $\mu(\theta)$  and  $\Sigma(\theta)$  will not be explicitly known but have to be estimated via simulation. Hence synthetic likelihoods can be viewed as an instance of the simulated method of moments, [McFadden \[1989\]](#).

The implementation of SL is straightforward. For a given  $\theta$  synthetic datasets  $y^{*1}, \dots, y^{*N}$  are generated independently from the model, hence each entry in vector  $y^{*n}$  belongs to the space  $\mathcal{Y}$  (see the notation at the beginning of section 3) and  $\dim(y^{*n}) = \dim(y)$ ,  $n = 1, \dots, N$ . Summary statistics  $s^{*n} = s(y^{*n})$  are computed for each simulated dataset and from these we obtain the moments estimates:

$$\hat{\mu}_N(\theta) = \frac{1}{N} \sum_{n=1}^N s^{*n}, \quad \hat{\Sigma}_N(\theta) = \frac{1}{N-1} \sum_{n=1}^N (s^{*n} - \hat{\mu}_N(\theta))(s^{*n} - \hat{\mu}_N(\theta))'.$$

Note that the only parameter that needs to be tuned in order to construct the synthetic likelihoods is  $N$ . When applying SL to SDEMEdMs it is important that summary statistics reflect the hierarchical structure of the model. To explain the intra- and inter-individual variance of the model we construct subject-specific summary statistics  $s_i := s^{\text{intra}}(y_i)$  for  $i = 1, \dots, M$ , as well as summaries that represent inter-individuals variation  $s^{\text{inter}} := s^{\text{inter}}(y_1, \dots, y_M)$ , so that  $s = (s_1, \dots, s_M, s^{\text{inter}})$ .

It is important to note that approximate normality for summary statistics can often be argued theoretically, e.g. by appealing to the central limit theorem (CLT). If the sample size is small or the summary statistics do not admit a CLT, then the normal assumption would have to be verified empirically using simulation. We refer to [Wood \[2010\]](#) for further details. We will not detail the computations involved in simulated methods of moments, but instead illustrate how SL applies to SDEMEdMs within a Bayesian framework. Here we follow [Price et al. \[2016\]](#) who proposed a fully Bayesian approach, sampling from the exact posterior  $\pi(\theta|s)$  without incurring into any bias caused by the choice of a finite  $N$ . The key feature exploits the idea underlying the pseudo-marginal method discussed in section 3, where an unbiased estimator is used in place of the unknown likelihood function. [Price et al. \[2016\]](#) note that plugging-in the estimates  $\hat{\mu}(\theta)$  and  $\hat{\Sigma}(\theta)$  into the Gaussian likelihood  $p(s|\theta)$  results in a biased estimate, while one could instead use the unbiased estimator of [Ghurye and Olkin \[1969\]](#) given by

$$\begin{aligned} \hat{p}(s|\theta) &= (2\pi)^{-d/2} \frac{c(d, N-2)}{c(d, N-1)(1-1/N)^{d/2}} |(N-1)\hat{\Sigma}_N(\theta)|^{-(n-d-2)/2} \\ &\times \left\{ \psi((N-1)\hat{\Sigma}_N(\theta) - (s - \hat{\mu}_N(\theta))(s - \hat{\mu}_N(\theta))'/(1-1/N)) \right\}^{(N-d-3)/2}. \end{aligned} \quad (10)$$

Here  $d = \dim(s)$ ,  $\pi$  denotes the mathematical constant (not the prior),  $N > d + 3$ , and for a square matrix  $A$  the function  $\psi(A)$  is defined as  $\psi(A) = |A|$  if  $A$  is positive definite and  $\psi(A) = 0$  otherwise. Finally  $|A|$  is the determinant of  $A$  and  $c(k, v) = 2^{-kv/2} \pi^{-k(k-1)/4} / \prod_{i=1}^k \Gamma(\frac{1}{2}(v-i+1))$ . Algorithm 4 is parallel to algorithm 1, but uses SL to draw from the posterior  $\pi(\theta|s)$  instead of  $\pi(\theta|y)$ .

For the implementation of algorithm 4, note that multiplicative constants such as the  $c(k, v)$ 's appearing in (10) are independent of  $\theta$  and cancel from the likelihood ratio

---

**Algorithm 4** Bayesian synthetic likelihoods

---

**Input:** a positive integer  $R$ . The observed summary statistics  $s$ . Fix a starting value  $\theta^*$  or generate it from the prior  $\pi(\theta)$ . Set  $\theta_1 = \theta^*$ . Choose a kernel  $q(\theta'|\theta)$ .

**Output:**  $R$  correlated samples from  $\pi(\theta|s)$ .

1. Conditionally on  $\theta^*$  generate independently  $N$  summaries  $s^{*1}, \dots, s^{*N}$ , compute moments  $\hat{\mu}_N(\theta^*)$ ,  $\hat{\Sigma}_N(\theta^*)$  and  $\hat{p}(s|\theta^*)$  from (10).
2. Generate a  $\theta^\# \sim q(\theta^\#|\theta^*)$ . Conditionally on  $\theta^\#$  generate independently  $s^{\#1}, \dots, s^{\#N}$ , compute  $\hat{\mu}_N(\theta^\#)$ ,  $\hat{\Sigma}_N(\theta^\#)$  and  $\hat{p}(s|\theta^\#)$ .
3. Generate a uniform random draw  $u \sim U(0, 1)$ , and calculate the acceptance probability

$$\alpha = \min \left[ 1, \frac{\hat{p}(s|\theta^\#)}{\hat{p}(s|\theta^*)} \times \frac{q(\theta^*|\theta^\#)}{q(\theta^\#|\theta^*)} \times \frac{\pi(\theta^\#)}{\pi(\theta^*)} \right].$$

If  $u > \alpha$ , set  $\theta_{r+1} := \theta_r$  otherwise set  $\theta_{r+1} := \theta^\#$ ,  $\theta^* := \theta^\#$  and  $\hat{p}(s|\theta^*) := \hat{p}(s|\theta^\#)$ . Set  $r := r + 1$  and go to step 4.

4. Repeat steps 2–3 as long as  $r \leq R$ .
- 

that defines the acceptance probability. To prevent the MCMC algorithm from reaching a premature halt, we suggest to set  $\hat{p}(s|\theta) := 0$  whenever the argument of  $\psi(\cdot)$  in (10) is not a positive definite matrix (except for the starting value  $\theta^*$ , of course). In the case study, section 6.2, we used algorithm 4 to estimate the model parameters of the SDEMEMs (4) and (5). To the best of our knowledge, this is the first application of the synthetic likelihood methodology to SDEMEMs. The methodology has also been tested with success in the simulation study in section 7.

## 5 Maximum likelihood estimation using SAEM

In this section we show how to construct maximum likelihood inference for the SDEMEMs (4) and (5) by using a special implementation of the stochastic approximation EM-algorithm (SAEM), originally proposed in [Delyon et al. \[1999\]](#). Actually, we also discuss the lack of reasonable results for our case study application, due to the difficulty of implementing SAEM for SDEMEMs.

Let  $y_i = (y_{i1}, \dots, y_{in_i})$  denote the  $i$ th longitudinal series of observation and set  $x_i = (x_{i0}, \dots, x_{in_i})$ . We further denote by  $x_{ij}^* = \log(V_{ij}^*)$  the log-transformed latent volumes for either of  $\star = \text{' '}$ ,  $\text{'surv'}$ , or  $\text{'kill'}$ . We denote by  $\Delta x_{ij}^* = x_{ij}^* - x_{i,j-1}^*$  its increments and  $\Delta t_{ij} = t_{ij} - t_{i,j-1}$ . Note that  $x_{i0} = \log(v_{i,0})$ , where we assume  $v_{i,0}$  deterministic, and that  $x_{ij}^* = \sum_{j'=1}^j \Delta x_{ij'}^*$  for  $j \geq 1$ . Let  $\eta$  denote the variance parameter(s) in the stochastic differential equation (4) or (5) and set  $\theta = (\bar{\mu}, \bar{\Sigma}, \eta, \sigma_\varepsilon^2)$ . In model (4), for unperturbed growth we have  $\phi_i = (\log(v_{i0}), \beta_i)$  and  $\eta = \gamma^2$ . In model (5), for response to treatment and regrowth we have  $\phi_i = (\log(v_{i0}), \alpha_i, \beta_i, \delta_i)$  and  $\eta = (\gamma^2, \tau^2)$ .

### 5.1 Calculations for the one-compartment model

Model (4) is one-compartment, hence no treatment is administered and notation-wise we have  $\star = \text{' '}$ . In this case  $\phi_i = \beta_i$ , however we use a more general notation as if  $\phi_i$  was vector-valued, which could be useful as a reference for further study. We assume that random effects are distributed as

$$\phi_i \sim \mathcal{N}(\bar{\mu}, \bar{\Sigma}),$$

where  $\bar{\Sigma}$  is diagonal; as mentioned above, in practice here we have  $\bar{\mu} \equiv \bar{\beta}$  and  $\bar{\Sigma} \equiv \sigma_\beta^2$ . The “complete likelihood” (using the terminology employed in [Dempster et al. 1977](#) when introducing the celebrated EM algorithm) can be expressed as

$$p(y, x, \phi | \theta) = \prod_{i=1}^M p(y_i | x_i, \phi_i, \sigma_\varepsilon^2) p(x_i | \eta, \phi_i) p(\phi_i | \bar{\mu}, \bar{\Sigma}). \quad (11)$$

For this model the solution process for the unobserved true volume is  $V_i(t) = v_{i,0} e^{\beta_i t + \gamma B_i(t)}$ , which implies that at sampling times we have  $V_i(t_j) = V_i(t_{j-1}) e^{\beta_i \Delta t_{ij} + \gamma \Delta B_{ij}}$  with  $\Delta t_{ij} = t_{ij} - t_{i,j-1}$ , then  $\Delta B_{ij} \sim \mathcal{N}(0, \Delta t_{ij})$  are independent increments of the Brownian motion, and finally by taking logarithms  $X_{ij} = X_{i,j-1} + \beta_i \Delta t_{ij} + \gamma \Delta B_{ij}$ . This means that the transition densities for subject  $i$  can be written  $p(x_{ij} | x_{i,j-1}) \equiv \mathcal{N}(x_{i,j-1} + \beta_i \Delta t_{ij}, \gamma^2 \Delta t_{ij})$ .

Due to the independent normally distributed increments of the Brownian motion, all of the conditional distributions entering the likelihood are normal and the complete log-likelihood function takes the form

$$\begin{aligned} l(y, x, \phi | \theta) &= c - \sum_{i=1}^M \left[ \frac{n_i}{2} \log(\sigma_\varepsilon^2) + \frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^{n_i} (y_{ij} - x_{ij})^2 \right] \\ &\quad - \sum_{i=1}^M \left[ \frac{n_i}{2} \log(\gamma^2) + \sum_{j=1}^{n_i} \frac{(\Delta x_{ij} - \beta_i \Delta t_{ij})^2}{2\gamma^2 \Delta t_{ij}} \right] \\ &\quad - \frac{M}{2} \log |\Sigma| - \frac{1}{2} \sum_{i=1}^M (\phi_i - \bar{\mu}) \Sigma^{-1} (\phi_i - \bar{\mu})' \end{aligned} \quad (12)$$

where  $c$  is a normalizing constant. In particular, this loglikelihood belongs to the *curved exponential family*, i.e. it can be written as

$$l(y, x, \phi | \theta) = -\Psi(\theta) + \langle S(y, x, \phi), \xi(\theta) \rangle$$

where  $\langle \cdot, \cdot \rangle$  denotes the scalar product,  $\Psi(\cdot)$  and  $\xi(\cdot)$  are functions of  $\theta$  and  $S(\cdot)$  is the so-called minimal sufficient statistic of the complete likelihood. Here the sufficient statistic is defined as  $S(y, x, \phi) = (S_{\bar{\mu}_\ell}, S_{\bar{\Sigma}_{\ell, \ell'}}, S_{\gamma^2}, S_{\sigma_\varepsilon^2})$ , for  $1 \leq \ell, \ell' \leq \dim(\phi)$ , with

$$\begin{aligned} S_{\sigma_\varepsilon^2} &= \sum_{i=1}^M \sum_{j=1}^{n_i} (y_{ij} - x_{ij})^2, & S_{\gamma^2} &= \sum_{i=1}^M \sum_{j=1}^{n_i} \frac{(\Delta x_{ij} - \beta_i \Delta t_{ij})^2}{\Delta t_{ij}} \\ S_{\bar{\mu}_\ell} &= \sum_{i=1}^M \phi_{\ell i}, & S_{\bar{\Sigma}_{\ell, \ell'}} &= \sum_{i=1}^M (\phi_{\ell i} - \bar{\mu}_\ell)(\phi_{\ell' i} - \bar{\mu}_{\ell'})'. \end{aligned}$$

The complete data maximum likelihood estimators are given by

$$\begin{aligned} \hat{\sigma}_\varepsilon^2 &= \frac{S_{\sigma_\varepsilon^2}}{\sum_{i=1}^M n_i}, & \hat{\gamma}^2 &= \frac{S_{\gamma^2}}{\sum_{i=1}^M n_i}, \\ \hat{\mu}_\ell &= \frac{S_{\bar{\mu}_\ell}}{M}, & \hat{\Sigma}_{\ell, \ell'} &= \frac{S_{\bar{\Sigma}_{\ell, \ell'}}}{M}. \end{aligned}$$

The SAEM algorithm consists of alternating three steps until convergence. Denote with  $k \geq 1$  a generic iteration of SAEM and with  $\hat{\theta}^{(0)}$  and  $S_0$  starting values for the parameters and for the sufficient statistics respectively:

- i: sample  $(x^{(k)}, \phi^{(k)})$  from  $p(x, \phi|y, \hat{\theta}^{(k-1)})$ .
- ii: Compute  $S^{(k)} = S^{(k-1)} + \omega_k \{S(y, x^{(k)}, \phi^{(k)}) - S^{(k-1)}\}$ .
- iii: Compute  $\hat{\theta}^{(k)} = \arg \max_{\theta} (-\Psi(\theta) + \langle S_k, \xi(\theta) \rangle)$ , then set  $k := k + 1$  and go to (i), unless convergence of  $\{\hat{\theta}^{(k)}\}$  is apparent.

Therefore at iteration  $k$  step (ii) updates  $S^{(k)} = (S_{\bar{\mu}_\ell}^{(k)}, S_{\bar{\Sigma}_{\ell, \ell'}}^{(k)}, S_{\gamma^2}^{(k)}, S_{\sigma_\varepsilon^2}^{(k)})$ , while step (iii) returns  $\hat{\theta}^{(k)} = (\hat{\mu}_\ell^{(k)}, \hat{\Sigma}_{\ell, \ell'}^{(k)}, \hat{\gamma}^{2(k)}, \hat{\sigma}_\varepsilon^{2(k)})$ . The sequence  $\{\omega_k\}_{k \geq 1}$  must be chosen so that  $0 \leq \omega_k \leq 1$ ,  $\sum_{k=1}^{\infty} \omega_k = \infty$ , and  $\sum_{k=1}^{\infty} \omega_k^2 < \infty$ . If this holds, then the sequence  $\{\hat{\theta}^{(k)}\}_{k \geq 1}$  converges to a stationary point of the (incomplete) data likelihood  $p(y; \theta)$  as  $k \rightarrow \infty$  under the regularity conditions of [Delyon et al. \[1999\]](#).

For more general models than the one under consideration, sampling in step (i) is non-trivial. In the Bayesian setting, an important advancement is the development of sequential Monte Carlo filters (particle filters), such as the particle MCMC (pMCMC) class of algorithms studied in [Andrieu et al. \[2010\]](#). In fact, for SDEMEmS [Donnet and Samson \[2014\]](#) use pMCMC to sample from  $p(x, \phi|y, \cdot)$ . In our case, since measurements  $y_i$  from different subjects are assumed independent, and since the same is assumed for the corresponding latent trajectories  $x_i$  and random effects  $\phi_i$ , we have that conditionally on a given  $y_i$ , the entries in  $x$  and  $\phi$  are independent, that is we have  $p(x, \phi|y, \cdot) = \prod_{i=1}^M p(x_i, \phi_i|y_i, \cdot)$ . Therefore at iteration  $k$  we can simulate  $(x_i^{(k)}, \phi_i^{(k)})$  from the corresponding  $p(x_i, \phi_i|y_i, \cdot)$  and finally form  $(x^{(k)}, \phi^{(k)}) = (x_{1:M}^{(k)}, \phi_{1:M}^{(k)})$ . In order to sample from each  $p(x_i, \phi_i|y_i, \cdot)$  we use a Metropolis-within-Gibbs procedure which we detail in appendix.

## 5.2 Calculations for the two-compartments model

In this case we have  $\phi_i = (\beta_i, \delta_i, \alpha_i)$ . The complete likelihood for model (5) is similar to (11), only  $p(x_i|\eta, \phi_i)$  must be replaced by the product of  $p(x_i^{\text{surv}}|\gamma^2, \phi_i)$  and  $p(x_i^{\text{kill}}|\tau^2, \phi_i)$ . For model (5) MLEs for  $\gamma^2$  and  $\tau^2$  are computed from  $x^{\text{surv}}$  and  $x^{\text{kill}}$  but otherwise expressions are similar. From the derivations performed in section 5.1 it is clear that we have  $p(x_{ij}^{\text{surv}}|x_{i,j-1}^{\text{surv}}) \equiv \mathcal{N}(x_{i,j-1}^{\text{surv}} + \beta_i \Delta t_{ij}, \gamma^2 \Delta t_{ij})$  with  $x_{i0}^{\text{surv}} = \log((1 - \alpha_i)v_{i0})$ . Then we have  $p(x_{ij}^{\text{kill}}|x_{i,j-1}^{\text{kill}}) \equiv \mathcal{N}(x_{i,j-1}^{\text{kill}} - \delta_i \Delta t_{ij}, \tau^2 \Delta t_{ij})$  with  $x_{i0}^{\text{kill}} = \log(\alpha_i v_{i0})$ . The procedure to determine sufficient statistics is completely analogous to the one-compartment model. Also note that regarding the estimation of sufficient statistics for  $\bar{\alpha}$  and  $\sigma_\alpha$ , this is not different from computing sufficient statistics for the moments of a Gaussian likelihood, even though the random effects  $\alpha_i$  are distributed according to a truncated Gaussian. This is because sufficient statistics for truncated location-scale distributions are unaffected by the (constant) truncation points. With these considerations, we have that the expressions for  $S_{\sigma_\varepsilon^2}$ ,  $S_{\bar{\mu}_\ell}$  and  $S_{\bar{\Sigma}_{\ell, \ell'}}$  are the same as in section 5.1, just as the corresponding maximizers from the M-step. In addition we have

$$S_{\gamma^2} = \sum_{i=1}^M \sum_{j=1}^{n_i} \frac{(\Delta x_{ij}^{\text{surv}} - \beta_i \Delta t_{ij})^2}{\Delta t_{ij}}, \quad S_{\tau^2} = \sum_{i=1}^M \sum_{j=1}^{n_i} \frac{(\Delta x_{ij}^{\text{kill}} + \delta_i \Delta t_{ij})^2}{\Delta t_{ij}}$$

and the corresponding maximizers from the M-step follow as for the one-compartment model,  $\hat{\gamma}^2 = S_{\gamma^2} / \sum_{i=1}^M n_i$ ,  $\hat{\tau}^2 = S_{\tau^2} / \sum_{i=1}^M n_i$ .

Now, since the dynamics for the surviving and killed cells are assumed independent as from model (5), in order to sample from the conditional distributions necessary to

implement Gibbs sampling, the idea is to sample separately the two volume components. This means sampling separately from the conditionals for  $v_{ij}^{\text{surv}}$  and  $v_{ij}^{\text{kill}}$ , then given these two samples form deterministically  $v_{ij} := v_{ij}^{\text{surv}} + v_{ij}^{\text{kill}}$  and therefore obtain a sample for  $x_{ij} := \log v_{ij}$ . Of course we also have  $x_{ij}^{\text{surv}} := \log v_{ij}^{\text{surv}}$  and  $x_{ij}^{\text{kill}} := \log v_{ij}^{\text{kill}}$ . The Gibbs sampling for  $v_{ij}^{\text{surv}}$  and  $v_{ij}^{\text{kill}}$  can be carried separately for the two volumes in the same way as performed for the one-compartment case.

## 6 Case study

We received data from a tumor xenography study including four treatment groups (1: chemo therapy, 2: radiation therapy, 3: combination therapy I, 4: combination therapy II), and one untreated control (group 5). Each group consists of 7-8 mice. Mice were followed up on Mondays, Wednesdays, and Fridays for six consecutive weeks or until their tumor volume exceeded 1,000 mm<sup>3</sup> in which case the mouse was sacrificed as prescribed by the ethical guidelines. In groups 2 and 4 about half of the mice were sacrificed within the treatment period or shortly after, and due to the reduced sample sizes these are not further considered.

Treatment in equal size doses was applied on days 1, 4, and 6 of the study. Afterwards no treatment was administered. The repeated measurements of tumor volumes in the three remaining groups are shown in Figure 1. It is obvious that growth patterns vary substantially between subjects. In the untreated control group a single mouse with a slowly growing tumor survived for 32 days before sacrifice while all other untreated mice got sacrificed within 10 days. In the active treatment groups we see patterns of decay followed by regrowth which match the characteristic shape of the double exponential curve. In the same groups we also see tumors that appear to grow continuously, unaffected by the treatment. An outlying mouse in group 1 displays a slowly vanishing tumor: for this mouse, most likely the tumor cells never started growing in the first place, so what is measured here is the thickness of the skin, and therefore data for this mouse are not considered in our analyses. Several mice display tumor volumes that are stable over shorter durations of time. These stable periods deviate from the growth patterns of the ordinary simple and double exponential mixed models, but can be explained by the random variations in growth and decay rates which are modeled in the double exponential SDEMEM.

We applied the double exponential SDEMEM (5) to model the log-tumor volumes post treatment (i.e. starting from day 6). Separate model fits were obtained from treatment groups 1 and 3. Tumor growth in the untreated controls was modeled with the simple exponential SDEMEM (4) starting from day 1 of the study.

For Bayesian analysis of the double exponential SDEMEM (5) we choose a truncated Gaussian prior on the average treatment effect  $\bar{\alpha} \sim N_{[0,1]}(0.6, 0.2^2)$  ( $N_{[a,b]}(\cdot)$  denotes a Gaussian truncated to  $[a, b]$ ) which assigns strictly positive probabilities to  $\bar{\alpha}$  values near zero and one. This is to anticipate that an effective treatment could have the effect that tumors are completely eliminated while an inefficient treatment might not kill any tumor cells. For the remaining parameters in (5) we choose the priors  $\log \bar{\beta} \sim \mathcal{N}(0.7, 0.6^2)$ ,  $\log \bar{\delta} \sim \mathcal{N}(0.7, 0.6^2)$ ,  $\sigma_{\beta} \sim \text{InvGam}(4, 2)$ ,  $\sigma_{\delta} \sim \text{InvGam}(4, 2)$ ,  $\sigma_{\alpha} \sim \text{InvGam}(5, 1.5)$ ,  $\gamma \sim \text{InvGam}(5, 7)$ ,  $\tau \sim \text{InvGam}(5, 7)$ , and  $\sigma_{\epsilon} \sim \text{InvGam}(2, 1)$ , where  $\text{InvGam}(\cdot)$  denotes the inverse-Gamma distribution. Note that positive model parameters have been reparametrised by their logarithms.



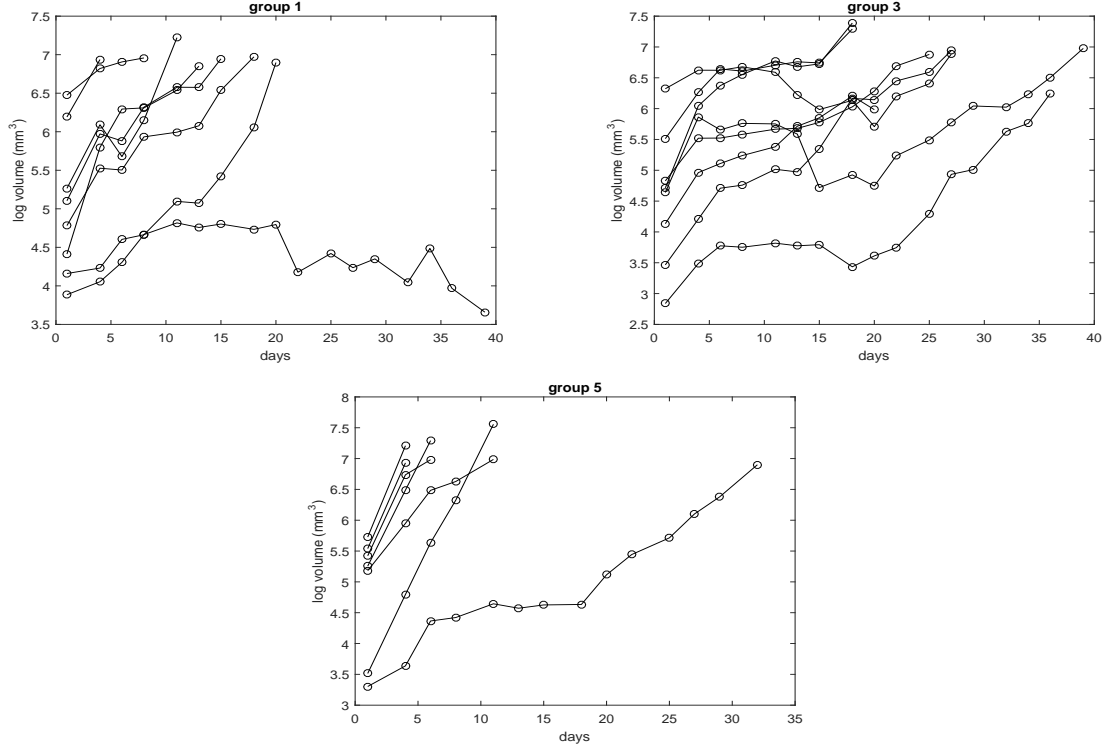


Figure 1: Data of log-volumes ( $mm^3$ ) for three treatment groups.

## 6.1 Results using exact Bayesian inference

We fitted model (5) separately for groups 1 and 3 using exact Bayesian inference as described in Section 3. For each subject, we set  $v_{i0}$  equal to the corresponding measured volume at day 1; however recall that, by looking at model (5), starting states depend also on  $\alpha_i$ . Algorithm 1 was started at  $\log \bar{\beta} = 1.6$ ,  $\log \bar{\delta} = 1.6$ ,  $\log \bar{\alpha} = -0.36$ ,  $\log \gamma = 0$ ,  $\log \tau = 0$ ,  $\log \sigma_{\beta} = -0.7$ ,  $\log \sigma_{\delta} = -0.7$ ,  $\log \sigma_{\alpha} = -2.3$ ,  $\log \sigma_{\varepsilon} = 0$ . We used  $L = 3,000$  particles to approximate the likelihood using algorithm 2 but resampling particles only when  $ESS < L/3$ . Chains of length  $R = 30,000$  were produced including a burn-in of 10,000 iterations. Computations took about 330 seconds for treatment group 1 ( $M = 5$ ), and 1,000 seconds for treatment group 3 ( $M = 8$ ) with a MATLAB code running on a Intel Core i7-4790 3.60 GHz. For both groups average acceptance rates observed during the execution of algorithm 1 were equal to 30%.

Results are shown in Table 1. Compared to group 1, it is evident that the larger sample size decreases the posterior variability considerably in group 3. The treatment efficacy is estimated at  $\bar{\alpha} = 37\%$  in group 3 and at  $\bar{\alpha} = 41\%$  in group 1. Unfortunately, this is unlikely to be a statistically significant difference, as the posterior for  $\bar{\alpha}$  is very wide in both groups; e.g. for group 3 see Figure 2, comparing the posterior distributions to the corresponding priors. In section 7.1 we show that having at disposal groups with a larger number of subjects enables the identification of the treatments efficacy  $\alpha$ . Note that posteriors for  $\log \bar{\beta}$ ,  $\gamma$ ,  $\tau$  and  $\sigma_{\varepsilon}$  are informative when compared to their priors. Also, the estimate for  $\bar{\beta}$  is higher in group 1 than in group 3, as it should be by looking at Figure 1, where it is clear that the tumor growth rate in group 1 is faster than in group 3 (recall in group 1 the decaying trajectory was not considered in the analysis). It is reassuring that the measurement error variance is estimated consistently by  $\hat{\sigma}_{\varepsilon} \simeq 0.1$  in all groups. This means that tumor volumes were measured with a relative accuracy approximately

Table 1: Posterior means and 95% posterior intervals: for each parameter we first report exact Bayesian inference using the particle marginal method and then approximate Bayesian inference using synthetic likelihoods estimation.

	group 1		group 3		group 5	
$\beta$	6.80	[4.64,9.03]	4.03	[2.88,5.29]	7.04	[4.97,9.00]
	6.49	[4.74,8.43]	3.67	[2.79,4.71]	8.78	[6.97,10.95]
$\bar{\delta}$	2.06	[0.61,5.28]	1.70	[0.60,3.46]	—	
	2.04	[0.53, 5.54]	1.50	[0.58,3.49]		
$\bar{\alpha}$	0.37	[0.04,0.77]	0.41	[0.13,0.74]	—	
	0.33	[0.03,0.67]	0.34	[0.04,0.70]		
$\gamma$	1.19	[0.74,1.75]	1.22	[0.90,1.53]	1.39	[1.06,1.82]
	1.05	[0.71,1.43]	0.92	[0.59,1.24]	1.27	[0.63,1.82]
$\tau$	1.41	[0.63,2.96]	2.22	[1.62,3.10]	—	
	1.35	[0.65,2.71]	1.75	[0.99,2.66]		
$\sigma_\beta$	0.50	[0.20,1.22]	0.46	[0.20,1.02]	0.50	[0.20,1.20]
	0.46	[0.20,1.08]	0.45	[0.21,0.89]	1.05	[0.24,3.07]
$\sigma_\delta$	0.45	[0.19,1.05]	0.43	[0.18,0.94]	—	
	0.49	[0.19,1.13]	0.45	[0.19,0.96]		
$\sigma_\alpha$	0.28	[0.13,0.54]	0.33	[0.14,0.69]	—	
	0.26	[0.13,0.51]	0.28	[0.13,0.54]		
$\sigma_\varepsilon$	0.11	[0.07,0.17]	0.07	[0.06,0.09]	0.15	[0.11,0.21]
	0.17	[0.10,0.27]	0.11	[0.08,0.17]	0.19	[0.11,0.30]

within  $\pm 20\%$  which is realistic in an experiment such as the one considered here. On the other hand, we found the marginal posterior for  $\sigma_\alpha$  to be highly sensitive to the choice of its prior; the posterior distribution followed the shape of the prior, regardless of the choice of hyperparameters.

The one-compartment model (4) was fitted to the untreated controls (group 5). The priors were the same as those for the corresponding parameters in the two-compartment model (5). Parameter estimates are shown in Table 1. Estimates of the mean population growth rate  $\bar{\beta}$  is again much higher than for group 3, as it should be, but very similar to group 1, which is again not surprising by looking at Figure 1. The diffusion coefficient  $\gamma$  is similar to the intensities of the stochasticity for the two-compartment model (as given by  $\gamma$  and  $\tau$ ) and the measurement error variance  $\sigma_\varepsilon$  is compatible with the previously fitted model, which is reassuring. The variance component  $\sigma_\beta$  was again difficult to identify, which is not surprising given the increased sparsity of the data for this group.

To make a rough assessment of whether model (5) is realistic compared to the data, we simulated three independent datasets using the posterior means in Table 1. The simulations are shown in Figures 3 and 4. The overall impression is that model (5) is capable of generating growth dynamics that are similar to the experimental data.

## 6.2 Results using synthetic likelihoods

We used  $N = 2,000$  simulations to construct the synthetic likelihoods approximation and performed Bayesian estimation using  $R = 20,000$  iterations of algorithm 4, as described in section 4.2. We used the same priors and initial parameter values as in the exact Bayesian analysis. During the execution of the algorithm we observed an acceptance rate of about 30% and the procedure required about 520 seconds for group 3 ( $M = 8$  subjects).

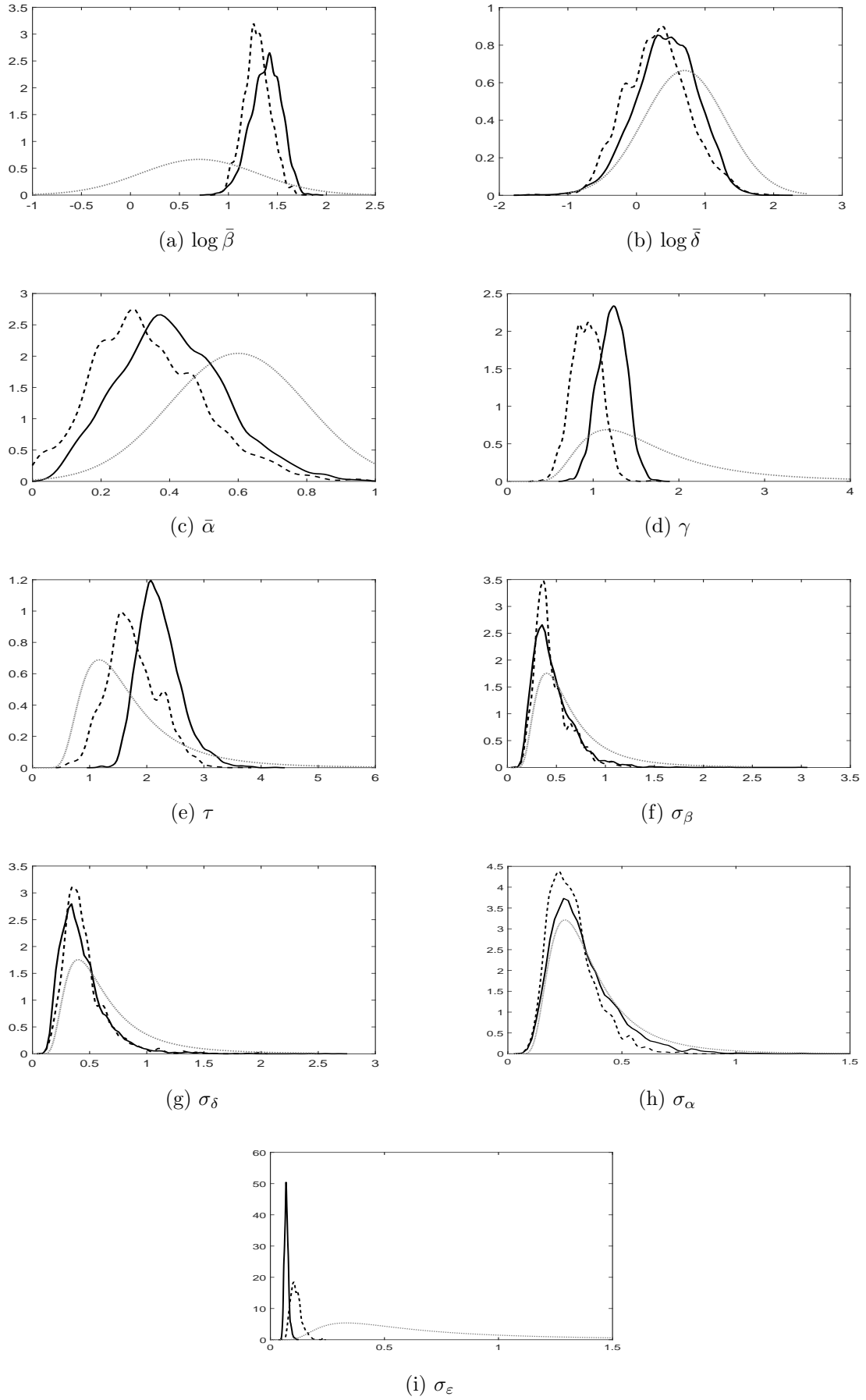


Figure 2: Treatment group 3, exact posteriors via particle marginal method (solid lines), synthetic likelihoods posteriors (dashed) and prior densities (dotted gray). The prior density for  $\sigma_{\epsilon}$  was multiplied by 4 for ease of comparison.

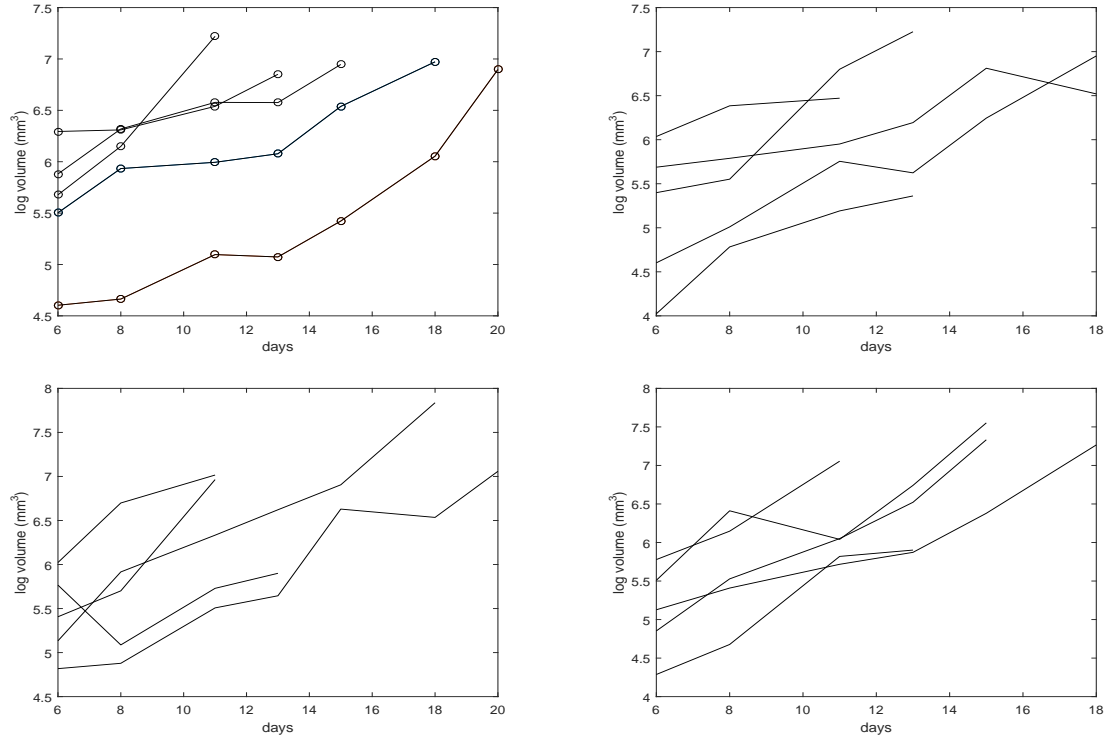


Figure 3: Fitted data in group 1 (top left) and three realizations from model (5) estimated with exact Bayesian methodology (remaining plots). Top left panel does not report data for one excluded mouse. Recall for this group measurements at days 1 and 4 were disregarded during estimation, hence times on abscissas start at day 6.

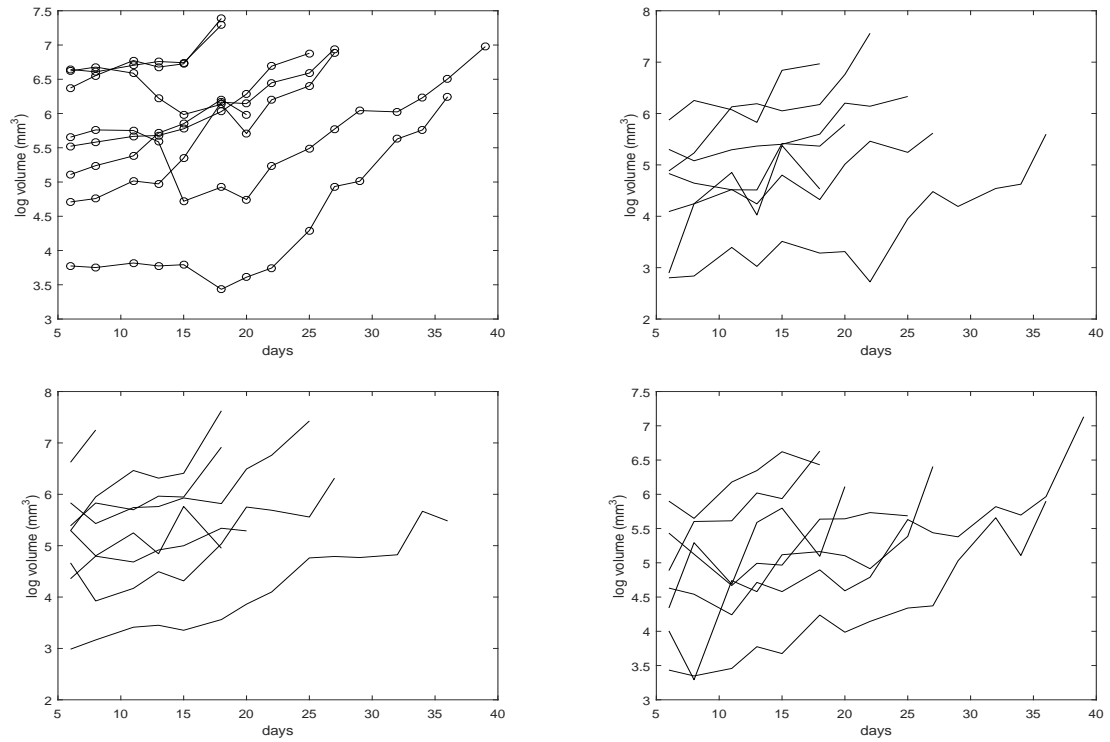


Figure 4: Fitted data in group 3 (top left) and three realizations from model (5) estimated with exact Bayesian methodology (remaining plots). Recall for this group measurements at days 1 and 4 were disregarded during estimation, hence times on abscissas start at day 6.

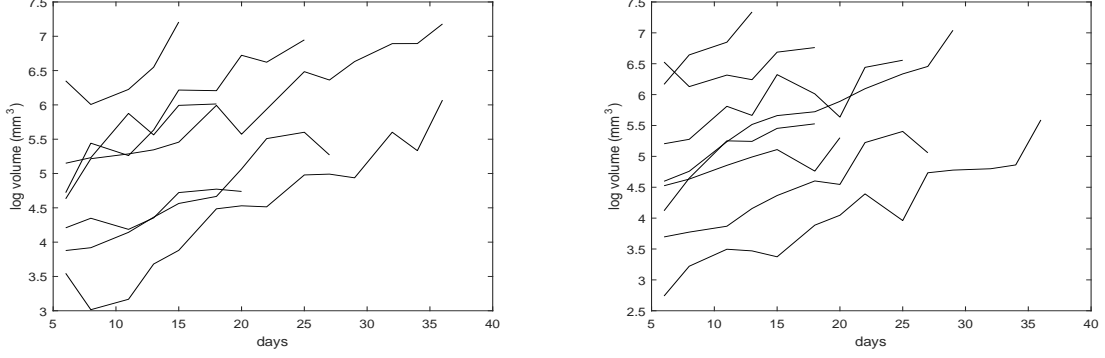


Figure 5: Group 3: two realizations from model (5) estimated with synthetic likelihoods.

Here we define the components for each individual (vector) statistic  $s_i := s^{\text{intra}}(y_i)$  as defined in section 4.2: (i) the mean absolute deviation for the repeated measurements  $\text{MAD}\{y_{ij}\}_{j=1:n_i}$  (ii) the slope of the line segment connecting the first and the last observation,  $(y_i(t_{n_i}) - y_i(t_1))/(t_{n_i} - t_1)$ ; (iii+iv) the values of the first and second measurements  $y_{i1}$  and  $y_{i2}$ ; (v) the estimated slope of a first order autoregressive fit of the repeated measurements, that is  $\hat{\beta}_{i1}$  from the regression  $E(y_{ij}) = \beta_{i0} + \beta_{i1}y_{i,j-1}$ . Note that when fitting model (4) to the control group, the last summary statistic was dropped to prevent  $\Sigma_N(\theta)$  from being singular (several mice had only two observations so that the second and fifth summary were perfectly correlated). Additional “population” inter-individuals summary statistics  $s^{\text{inter}}$  included: (i)  $\text{MAD}\{y_{i1}\}_{i=1:M}$ , the mean absolute deviation between subjects at the first time point (day 6 for the active treatment groups and day 1 for the control group); (ii) the same as in (i) but for the second time point.

Parameter estimates are presented in Table 1 and the approximate marginal posterior distributions for group 3 are shown in Figure 2. For those posteriors resulting different from the ones obtained with the exact Bayesian methodology (e.g.  $\gamma$  and  $\tau$  in Figure 2), using the corresponding posterior means to produce artificial trajectories does not seem to result in different behaviours: for example Figure 5 (using posterior means estimated using the synthetic likelihoods approach) shows trajectories similar to Figure 4. This suggests that synthetic likelihoods can return reasonable inference for SDEMEdMs, if the summary statistics are well chosen (see also the simulation study in section 7.1). Finally, Figure 6 gives normal qq-plots of simulated summaries corresponding to the last draw generated with algorithm 4: therefore, we report both the two summaries corresponding to inter-individual variation, as well as the five summaries for the intra-subject variation. All summaries appear fairly close to normality, however in all subjects  $s_1^{\text{intra}}$  and  $s_5^{\text{intra}}$  were those having tails slightly departing from the normality assumption.

### 6.3 Results for maximum likelihood inference using SAEM

We implemented  $K = 25,000$  iterations of SAEM, using coefficients  $\omega_k = 1$  for  $k \leq 20,000$  and  $\omega_k = (k - 20,000)^{-0.9}$  for  $20,000 < k \leq K$ . For both the one-compartment model (4) and the two-compartment model (5) we must specify variances for the Gaussian increments used to propose random parameters in the Metropolis-Hastings step, that is to propose  $\beta_i$  for model (4) and  $(\beta_i, \delta_i, \alpha_i)$  for model (5). The complexity of the setup, as outlined in the appendix, where random parameters  $\alpha_i$  also enter the states initial condition  $x_{i0}$  for model (5), coupled with the setup for the variances of the Gaussian

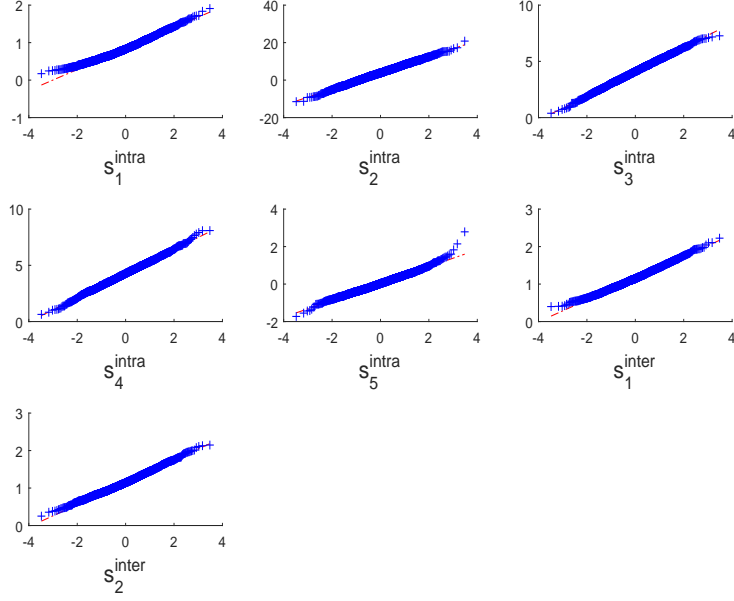


Figure 6: Group 3: normal qq-plots for the intra-individual summary statistics generated for a specific subject from group 3 ( $s_1^{\text{intra}}, \dots, s_5^{\text{intra}}$ ) as well as inter-individual summaries ( $s_1^{\text{inter}}, s_2^{\text{inter}}$ ). All summaries have been generated in correspondence of the last simulated parameter draw in the MCMC.

increments in Metropolis random walk, implies that SAEM can be quite sensitive to the chosen settings. In fact, at each iteration  $k$  of SAEM, the evolution of the algorithm is based on the sampling of a single trajectory for  $\{X_{ij}\}$  and the random effects. A preferable strategy would be to run particle MCMC, as in [Donnet and Samson \[2014\]](#), as in that case a swarm of trajectories is generated (both for the states and the random parameters) and the unlikely ones get down-weighted, so that the calculation of the likelihood is carried only by the most plausible trajectories, for each iteration of an MCMC procedure. Therefore the results of our SAEM simulations are unsatisfactory and not returning reasonable results.

## 7 Simulation studies

We first run a simulation study consisting in analysing thirty datasets generated independently from model (5), with ground-truth parameters set to the posterior means previously obtained using exact Bayesian methodology (particle marginal method, PMM) on group 3, as found in Table 1. Each of the thirty datasets has measurements for  $M = 8$  subjects, with observations generated at the same sampling times as for subjects in group 3, and using the same values  $x_{0i}$  as set for group 3 in the previous sections. For each of the thirty datasets we apply both the exact Bayesian and synthetic likelihoods procedure, initializing the corresponding algorithms at the same starting parameters as in previous analyses, namely  $\log \bar{\beta} = 1.6$ ,  $\log \bar{\delta} = 1.6$ ,  $\log \bar{\alpha} = -0.36$ ,  $\log \gamma = 0$ ,  $\log \tau = 0$ ,  $\log \sigma_{\beta} = -0.7$ ,  $\log \sigma_{\delta} = -0.7$ ,  $\log \sigma_{\alpha} = -2.3$ ,  $\log \sigma_{\varepsilon} = 0$ . We used exactly the same setup as from the case-study to estimate parameters using the two inference strategies, where the particle marginal method uses  $L = 3,000$  particles and  $R = 30,000$  MCMC iterations and the Bayesian synthetic likelihoods approach uses  $N = 2,000$  and  $R = 20,000$ .



Table 2: Simulation study: true parameter values ( $\theta_0$ ), median bias and RSME using the particle marginal MCMC method (PMM) and Bayesian synthetic likelihoods (BSL).

	$\bar{\beta}$	$\bar{\delta}$	$\bar{\alpha}$	$\gamma$	$\tau$	$\sigma_\beta$	$\sigma_\delta$	$\sigma_\alpha$	$\sigma_\epsilon$
$\theta_0$	4.03	1.70	0.41	1.22	2.22	0.46	0.43	0.33	0.07
PMM									
bias	-0.122	-0.044	0.258	-0.556	-0.675	-0.069	0.001	-0.125	0.120
RMSE	0.987	0.829	0.244	0.546	0.819	0.142	0.170	0.133	0.136
BSL									
bias	-0.299	-0.056	-0.133	-0.476	-1.020	-0.077	0.008	-0.102	0.231
RMSE	0.617	0.251	0.145	0.489	1.010	0.079	0.029	0.106	0.230

For both inference methods, we collected the corresponding thirty posterior means  $\hat{\theta}_b$  (after a burning of 10,000 iterations for the synthetic likelihoods experiments and 15,000 iterations for PMM). Also, by denoting with  $\theta_0$  the ground truth parameters, for the  $B = 30$  experiments we compute the median bias, that is the median of the  $B$  differences  $(\hat{\theta}_b - \theta_0)$  and the root mean square error (RMSE)  $\sqrt{\sum_{b=1}^B (\hat{\theta}_b - \theta_0)^2 / B}$ . When looking at the PMM results in Table 2 we notice that even in a simulation study the “data poor” scenario is challenging for SDEMEds, given the multiple sources of random variation considered for this class of models. The small bias for the drift parameters  $(\bar{\beta}, \bar{\delta})$  implies that the main dynamics are well identified, however it is more difficult to conclude something for other parameters, such as  $\bar{\alpha}$ ,  $\gamma$  and  $\tau$ . However, instead of looking at the analytic measures in Table 2, a more revealing view over the results is in Figure 7 where we notice that for most key parameters the variability of the bias is smaller for the synthetic likelihoods approach. Of course this does not mean that an approximate methodology should in general be preferred to exact inference, but it shows that in presence of multiple sources of variability and scarcity of data, the use of summary statistics can protect the inference against too noisy realizations of the model simulator. The latter intuition was a motivating factor for the work in Wood [2010]. Of course, both for PMM and Bayesian synthetic likelihoods it would be interesting to see results using a larger number of experiments  $B$ ; unfortunately this type of simulation study is computationally very intensive.

## 7.1 Results using larger simulated datasets

Here we run a further simulation study. We consider the problem of treatment identifiability assuming the availability of a larger group of subjects, i.e. a larger value for  $M$ . We simulate two sets of data, each having  $M = 17$  subjects. The two sets are simulated to highlight the role of the parameter  $\alpha$ , representing the treatment efficacy, see e.g. section 2.1. We produce a first dataset  $\mathcal{D}_1$ , this one emulating a group where the treatment has a low efficacy. We produce  $\mathcal{D}_1$  by setting parameters to the same values as for the PMM estimates for group 1 in Table 1, i.e. here  $\alpha = 0.37$ . The second dataset  $\mathcal{D}_2$  emulates a group where the treatment has higher efficacy, by setting  $\alpha = 0.75$  and the remaining parameters are equal to the PMM estimates for group 3 in Table 1. For both  $\mathcal{D}_1$  and  $\mathcal{D}_2$  we use Bayesian synthetic likelihoods to estimate parameters, with  $N = 6,000$ ,  $R = 15,000$  MCMC iterations and the same starting parameter values as in previous experiments. Results show that for both simulated datasets the corresponding ground-truth values for  $\alpha$

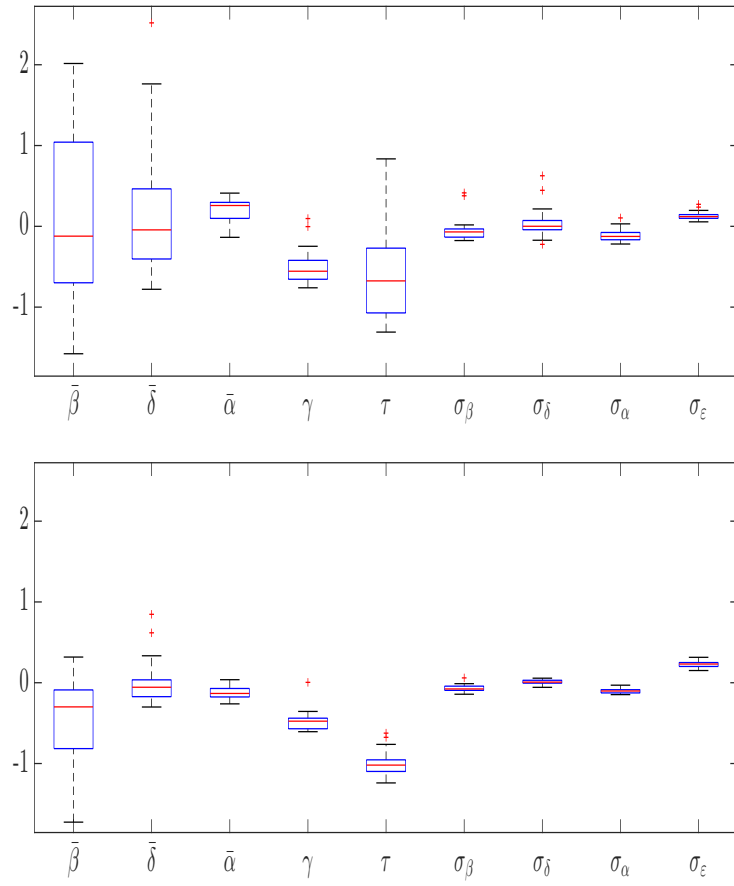


Figure 7: Simulation study with  $M = 8$  subjects: top, boxplots of the parameter estimates bias obtained with PMM; bottom, boxplots of the bias when using Bayesian synthetic likelihoods.

are correctly identified, see Figure 8, with marginal posteriors resulting more informative than for the real data experiments. This suggests that we could obtain more accurate inferences for treatments efficacy in real data, should larger experiments be conducted. The parameter  $\bar{\beta}$ , representing the tumor growth rate that should be higher for  $\mathcal{D}_1$  than for  $\mathcal{D}_2$ , indeed in Figure 8 it is shown to have the expected behaviour, and the two posteriors for  $\bar{\beta}$  are well separated. Diffusion coefficients are still difficult to identify, but this is a common characteristics in inference for diffusions in presence of data-poor scenarios. We can also notice that the residual variability  $\sigma_\epsilon$  is challenging to identify. When comparing the inference for  $\sigma_\epsilon$  with the real data results in Figure 2, we conjecture that for the real data experiment (where  $M$  is small) the inference injects considerable uncertainty into the estimate of  $\alpha$ . We therefore deduce that for small  $M$  we underestimate the residual variability, as a non-negligible fraction of the total variability is assigned to the treatment variability. For the simulated data with  $M = 17$  we have the opposite situation, where  $\alpha$  is estimated more precisely, this producing an enlarged residual variability  $\sigma_\epsilon$ . However the true value for the latter is not well captured, therefore it seems that we cannot obtain precise estimates simultaneously for both parameters.

## 8 Summary

We have introduced a new mixed effects model for the analysis of repeated measurements of tumor volumes in mice in a tumor xenograpy study. For each subject the dynamics for the exact, unobservable, tumor volumes are modeled by stochastic differential equations (SDEs), while observed volumes are assumed subject to measurement error. The resulting model is called a stochastic differential mixed effects model (SDEMEm), which is of state-space type. SDEMEmS provide a powerful representation for repeated measurement experiments, in that are able to explain the several sources of variability, namely: intra-individual stochastic variability; variability between subjects; residual variability (measurement error). We considered two different SDEMEmS, one for unperturbed growth modelling an untreated control group, and one for tumor (re)growth following an active treatment such as chemo- or radiation therapy. The former is a one-compartment model while the latter is a two compartments model. The two compartments represent the fraction of tumor cells that has been killed by the treatment and the fraction that has survived the treatment, respectively. Hence the model extends the classical double exponential model by including random perturbations in the growth dynamics.

Parameters inference for SDEMEmS is challenging due to their intractable likelihood functions. We have considered methods for exact and approximate inference, and in particular compared approximate Bayesian inference using the synthetic likelihoods (SL) approach (Wood, 2010, Price et al., 2016) to exact Bayesian inference based on a marginal particle method (Andrieu and Roberts, 2009). SL bases the inference on the likelihood function of normally distributed summary statistics, instead of the intractable likelihood function of the actual data. The efficiency of the resulting estimator relies on the choice of the summary statistics. For the application to SDEMEmS we advocated the use of subject specific summary statistics, which can be further comprised over groups by taking averages and computing covariances. In a case-study application to a tumor xenograpy study with four active treatment groups and an untreated control, we found that synthetic likelihoods performed similarly to exact Bayesian inference, indicating that our choice of summary statistics was appropriate. In spite of tiny sample sizes of 5 to 8 mice per group, we were able to estimate the population median tumor growth rate consistently across groups. Also

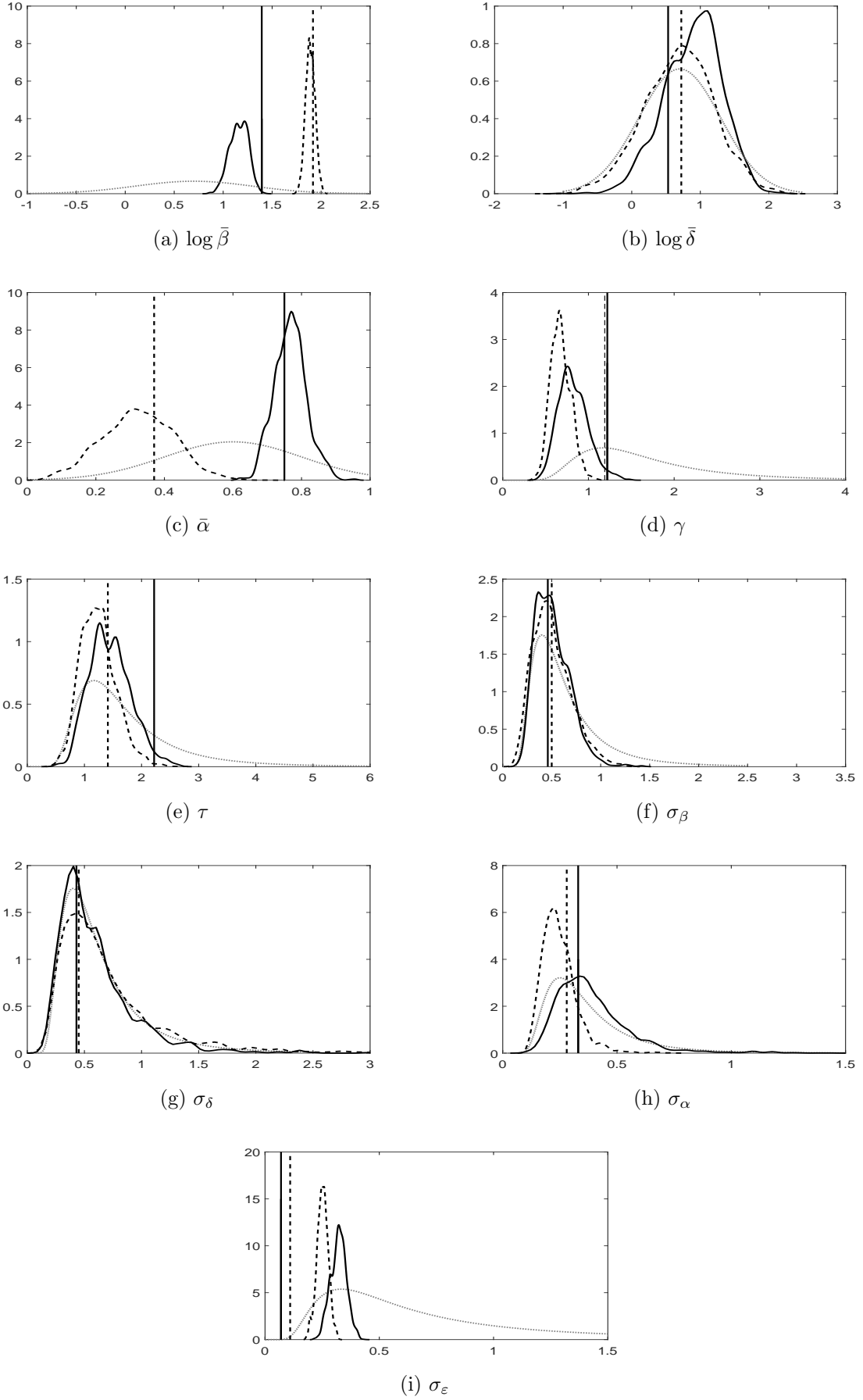


Figure 8: Posteriors based on simulated data  $\mathcal{D}_1$  (dashed curves) and based on simulated data  $\mathcal{D}_2$  (solid lines) obtained with Bayesian synthetic likelihoods. Dashed vertical lines indicate ground-truth parameters used for  $\mathcal{D}_1$ , while solid vertical lines indicate ground-truth parameters used for  $\mathcal{D}_2$ . Dotted lines denote prior densities.

the measurements error variance was accurately and consistently estimated across all the groups. Larger sample sizes are, however, needed to obtain accurate estimates of the treatment effects and their variance components. An advantage of synthetic likelihoods is that, unlike exact particle-based inference, it can be applied to models other than the state-space type. To the best of our knowledge ours is the first application of the synthetic likelihoods methodology to SDEMEmS.

Finally, a simulation study using artificial data showed that inference based on summaries of measurements, as advocated by the synthetic likelihoods approach, can result in robust inference against the many factors of variability considered in SDEMEmS. In comparison, exact Bayesian inference based on the full data resulted in posterior means having a larger bias.

## Acknowledgements

Research was partially supported by the Swedish Research Council (VR grant 2013-05167). We would like to thank the research team at the Center for Nanomedicine and Theranostics, DTU Nanotech, Denmark for providing the data for the case study and for introducing us to the problem of making inference from tumor xenography experiments.

## Appendix

We construct the conditional distribution enabling Gibbs sampling (in practice Metropolis-within-Gibbs) for the one-compartment model in section 5.1. For simplicity of notation we remove the conditioning on  $\theta$ , and we also assume that  $x_{i0}$  is deterministic and known. We write

$$p(x_i, \phi_i | y_i) = p(x_{i0}, x_{i1}, \dots, x_{i,n_i}, \phi_i | y_{i1}, \dots, y_{i,n_i}) \propto p(x_{i1}, \dots, x_{i,n_i}, \phi_i | y_{i1}, \dots, y_{i,n_i})$$

and in order to implement a single “sweep” of Gibbs sampling we need to execute the following two steps:

- a: only at  $j = 1$ , sample  $\phi_i \sim p(\phi_i | x_i, y_i)$ ;
- b: for  $j \geq 1$ , sample  $x_{ij} \sim p(x_{ij} | x_{i,0}, \dots, x_{i,j-1}, x_{i,j+1}, \dots, x_{i,n_i}, y_i, \phi_i) \equiv p(x_{ij} | x_{i,j-1}, x_{i,j+1}, y_{i,j}, \phi_i)$ .

The reason why in step (a)  $\phi_i$  is sampled only when  $j = 1$  is that  $\phi_i$  is time-invariant. The derivation of (b) is as follows: recall that  $y_i = (y_{i1}, \dots, y_{i,n_i})$ , then (b) can be written

$$p(x_{ij} | x_{i,0}, \dots, x_{i,j-1}, x_{i,j+1}, \dots, x_{i,n_i}, y_i, \phi_i) = p(x_{ij} | x_{i,0:j-1}, x_{i,j+1:n_i}, y_{i,1:j-1}, y_{i,j:n_i}, \phi_i).$$

Then, given  $x_{j-1}$  the latent process is independent on anything that happened before  $t_{j-1}$ , hence

$$p(x_{ij} | x_{i,0:j-1}, x_{i,j+1:n_i}, y_{i,1:j-1}, y_{i,j:n_i}, \phi_i) = p(x_{ij} | x_{i,j-1}, x_{i,j+1}, y_{i,j}, \phi_i).$$

In order to perform the sampling, we first need to write the conditional distributions in (a)-(b), starting from the corresponding joint density, which is (see (12) for reference)

$$\begin{aligned} p(x_i, y_i, \phi_i) &= p(y_i | x_i) p(x_i | \phi_i) p(\phi_i) = p(\phi_i) \prod_{j=1}^{n_i} p(y_{ij} | x_{ij}) p(x_{ij} | x_{i,j-1}; \phi_i) \\ &\propto \frac{1}{\sigma_\beta} e^{-\frac{1}{2\sigma_\beta^2}(\beta_i - \bar{\beta})^2} \frac{1}{\sigma_\varepsilon^{n_i}} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^{n_i} (y_{ij} - x_{ij})^2} \frac{1}{\gamma^{n_i}} e^{-\frac{1}{2\gamma^2} \sum_{j=1}^{n_i} \left[ \frac{(x_{ij} - x_{i,j-1} - \beta_i \Delta t_{ij})^2}{\Delta t_{ij}} \right]}. \end{aligned} \quad (13)$$

By looking at (13) we deduce the following

$$p(\phi_i|x_i, y_i) \propto e^{-\frac{1}{2\sigma_\beta^2}(\beta_i - \bar{\beta})^2} e^{-\frac{1}{2\gamma^2} \sum_{j=1}^{n_i} \left[ \frac{(x_{ij} - x_{i,j-1} - \beta_i \Delta t_{ij})^2}{\Delta t_{ij}} \right]} \quad (14)$$

and we can sample a single draw from (14) using a Metropolis-Hastings (MH) step. A MH step for (14) coupled with the Gibbs step (a) means that we run a so-called hybrid (Metropolis-within-Gibbs) Monte Carlo algorithm [Robert and Casella, 2013]. In practice if at a generic iteration of MH we have accepted a value  $\beta'_i$ , at next iteration a new value  $\beta_i^*$  is proposed from a kernel  $\beta_i^* \sim q(\beta_i^*|\beta'_i)$ , and we choose  $q(\cdot)$  to induce a Gaussian random walk, that is  $\beta_i^* = \beta' + r_\beta \cdot \xi$ , with  $\xi \sim \mathcal{N}(0, 1)$  and  $r_\beta$  the standard deviation of the Gaussian increments.. Regarding (b) we have that by adding and subtracting  $x_{i,j+1}$  in the third exponent in (13) we have, for  $j < n$

$$p(x_{ij}|x_{i,j-1}, x_{i,j+1}, y_{i,j}, \phi_i) \propto \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} (x_{ij} - y_{ij})^2 - \frac{1}{2\gamma^2 \Delta t_{ij}} [x_{ij}^2 - 2x_{ij}x_{i,j+1} + 2x_{ij}(x_{i,j+1} - x_{i,j-1}) - \beta_i \Delta t_{ij} x_{ij}] \right\}$$

while for  $j = n$

$$p(x_{ij}|x_{i,j-1}, x_{i,j+1}, y_{i,j}, \phi_i) \propto \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} (x_{ij} - y_{ij})^2 - \frac{1}{2\gamma^2 \Delta t_{ij}} (x_{ij} - x_{i,j-1} - \beta_i \Delta t_{ij})^2 \right\}.$$

Also  $p(x_{ij}|x_{i,j-1}, x_{i,j+1}, y_{i,j}, \phi_i)$  can be sampled via MH-within-Gibbs. But in this case we choose as kernel for proposing a new  $x_{ij}^*$  the transition density of  $\{X_t\}$ , that is  $x_{ij}^* \sim \mathcal{N}(x'_{i,j-1} + \beta_i \delta t_{ij}, \gamma^2 \Delta t_{ij})$ , with  $x'_{i,j-1}$  a value accepted for  $x_{i,j-1}$ . Hence to sample  $x_{ij}^*$  we don't need to introduce a random walk.

## References

- C. Andrieu and G. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, pages 697–725, 2009.
- C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society: Series B*, 72(3):269–342, 2010.
- M. Beaumont. Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164(3):1139–1160, 2003.
- O. Cappé, E. Moulines, and T. Rydén. *Inference in hidden Markov models*. Springer Science & Business Media, 2006.
- P. Del Moral. *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. New York: Springer, 2004.
- M. Delattre and M. Lavielle. Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Statistics and its interface*, 6(4):519–532, 2013.



- B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics*, pages 94–128, 1999.
- E. Demidenko. The assessment of tumour response to treatment. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(3):365–377, 2006.
- E. Demidenko. Three endpoints of in vivo tumour radiobiology and their statistical estimation. *International journal of radiation biology*, 86(2):164–173, 2010.
- E. Demidenko. *Mixed models: theory and applications with R*. John Wiley & Sons, 2013.
- A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B*, pages 1–38, 1977.
- S. Donnet and A. Samson. A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models. *Advanced Drug Delivery Reviews*, 65(7):929–939, 2013.
- S. Donnet and A. Samson. Using PMCMC in EM algorithm for stochastic mixed models: theoretical and practical issues. *Journal de la Société Française de Statistique*, 155(1):49–72, 2014.
- S. Donnet, J. Foulley, and A. Samson. Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics*, 66(3):733–741, 2010.
- A. Doucet, M. Pitt, G. Deligiannidis, and R. Kohn. Efficient implementation of Markov chain Monte Carlo when using an unbiased likelihood estimator. *Biometrika*, 2015. doi:doi:10.1093/biomet/asu075.
- C. Fuchs. *Inference for Diffusion Processes: With Applications in Life Sciences*. Springer Science & Business Media, 2013.
- S. Ghurye and I. Olkin. Unbiased estimation of some multivariate probability densities and related functions. *The Annals of Mathematical Statistics*, pages 1261–1271, 1969.
- N. Gordon, D. Salmond, and A. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *Radar and Signal Processing, IEE Proceedings F*, volume 140, pages 107–113, 1993.
- H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, pages 223–242, 2001.
- D. Heitjan. Generalized Norton-Simon models of tumour growth. *Statistics in medicine*, 10(7):1075–1088, 1991.
- D. Heitjan, A. Manni, and R. Santen. Statistical analysis of in vivo tumor growth experiments. *Cancer Research*, 53(24):6042–6050, 1993.
- G. Kitagawa. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of computational and graphical statistics*, 5(1):1–25, 1996.
- M. Kong and J. Yan. Modeling and testing treated tumor growth using cubic smoothing splines. *Biometrical Journal*, 53(4):595–613, 2011.

- T. Laajala, J. Corander, N. Saarinen, K. Mäkelä, S. Savolainen, M. Suominen, E. Alhoniemi, S. Mäkelä, M. Poutanen, and T. Aittokallio. Improved statistical modeling of tumor growth and treatment effect in preclinical animal studies with highly heterogeneous responses in vivo. *Clinical Cancer Research*, 18(16):4385–4396, 2012.
- J. Liu. Metropolized independent sampling with comparisons to rejection sampling and importance sampling. *Statistics and Computing*, 6(2):113–119, 1996.
- J. Marin, P. Pudlo, C. Robert, and R. Ryder. Approximate Bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180, 2012.
- P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26):15324–15328, 2003.
- D. McFadden. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5):995–1026, 1989.
- U. Picchini. Likelihood-free stochastic approximation EM for inference in complex models. 2016. [arXiv:1609.03508](https://arxiv.org/abs/1609.03508).
- M. Pitt, R. dos Santos Silva, P. Giordani, and R. Kohn. On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics*, 171(2):134–151, 2012.
- D. Prangle. Adapting the ABC distance function. [arxiv:1507.00874](https://arxiv.org/abs/1507.00874), 2015.
- L. Price, C. Drovandi, A. Lee, and D. Nott. Bayesian synthetic likelihood. 2016. <http://eprints.qut.edu.au/92795/>.
- C. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Science & Business Media, 2013.
- C. Sherlock, A. Thiery, G. Roberts, and J. Rosenthal. On the efficiency of pseudo-marginal random walk Metropolis algorithms. *The Annals of Statistics*, 43(1):238–275, 2015.
- S. Sisson and Y. Fan. *Handbook of Markov chain Monte Carlo*, chapter Likelihood-free MCMC. CRC Press, 2011.
- M. Stuschke, V. Budach, M. Bamberg, and W. Budach. Methods for analysis of censored tumor growth delay data. *Radiation research*, 122(2):172–180, 1990.
- T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31):187–202, 2009.
- G. Whitaker, A. Golightly, R. Boys, and C. Sherlock. Bayesian inference for diffusion driven mixed-effects models. *Bayesian Analysis*, 2016.
- S. Wood. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(7310):1102–1104, 2010.
- J. Wu. Confidence intervals for the difference of median failure times applied to censored tumor growth delay data. *Statistics in Biopharmaceutical Research*, 3(3):488–496, 2011.

- J. Wu and P. Houghton. Assessing cytotoxic treatment effects in preclinical tumor xenograft models. *Journal of biopharmaceutical statistics*, 19(5):755–762, 2009.
- C. Xia, J. Wu, and H. Liang. Model tumor pattern and compare treatment effects using semiparametric linear mixed-effects models. *Journal of Biometrics & Biostatistics*, 2013, 2013.
- L. Zhao, M. Morgan, L. Parsels, J. Maybaum, T. Lawrence, and D. Normolle. Bayesian hierarchical changepoint methods in modeling the tumor growth profiles in xenograft experiments. *Clinical Cancer Research*, 17(5):1057–1064, 2011.